#16 attachment
09/701586

UNITED STATES PATENT AND TRADEMARK OFFICE

I, Alan John SPARROW

translator to RWS Group plc, of Europa House, Marsham Way, Gerrards Cross, Buckinghamshire, England declare;

1. That I am a citizen of the United Kingdom of Great Britain and Northern Ireland.

2. That I am well acquainted with the German and English languages.

3. That the attached is, to the best of my knowledge and belief, a true translation into the English language of the accompanying copy of the specification filed with the application for a patent in Germany on 5 June 1998 under the number 198 25 213.7 and the official certificate attached hereto.

4. That I believe that all statements made herein of my own knowledge are true and that all statements made on information and belief are true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the patent application in the United States of America or any patent issuing thereon.

AJ Sparrow

For and on behalf of RWS Group plc

The 27th day of March 2003

FEDERAL REPUBLIC OF GERMANY

Certificate

BASF Aktiengesellschaft

of

Ludwigshafen/Germany

have filed a Patent Application under the title:

"Novel poly(ADP-ribose) polymerase genes"

on 5 June 1998 at the German Patent and Trademark Office.

The attached documents are a correct and accurate reproduction of the original submission for this Patent Application.

The German Patent and Trademark Office has for the time being given the Application the symbols C 12 N, C 07 H and A 01 K of the International Patent Classification.

Munich, 18 June 1999

German Patent and Trademark Office

The President

pp

Ebert

File No: 198 25 213.7

Novel poly(ADP-ribose) polymerase genes

The present invention relates to novel poly(ADP-ribose) polymerase (PARP) genes and to the proteins derived therefrom; antibodies with specificity for the novel proteins; pharmaceutical and gene therapy compositions which comprise products according to the invention; methods for the analytical determination of the proteins and nucleic acids according to the invention; methods for identifying effectors or binding partners of the proteins according to the invention, and methods for determining the activity of such effectors.

The primary physiological function of PARP (EC 2.4.2.30) (sometimes also referred to PARS, poly(adenosine-5'-diphosphoribose) synthetase) appears to be its involvement in a complex repair mechanism which cells have developed to repair DNA strand breaks. The primary cellular response to a DNA strand break appears moreover to consist of PARP-catalyzed synthesis of poly(ADP-ribose) from NAD^+ (cf. De Murcia, G. et al. (1994) TIBS, 19, 172).

PARP has a modular molecular structure. Three main functional elements have been identified to date: an N-terminal 46kDa DNA binding domain; a central 22kDa automodification domain to which poly(ADP-ribose) becomes attached, with the DNA affinity decreasing with increasing elongation; and a C-terminal 54 kDa NAD^+ binding domain. A leucine zipper region has been found within the automodification domain, indicating possible protein-protein interactions, only in the PARP from Drosophila. All PARPs known to date are active only as homodimers.

The high degree of organization of the molecule is reflected again in the strong conservation of the amino acid sequence. Thus, 62% conservation of the amino acid sequence has been found for PARP from humans, mice, cattle and chickens. There are greater structural differences from the PARP from Drosophila. The individual domains themselves in turn have clusters of increased coservation [sic]. Thus, the DNA binding region contains two so-called zinc fingers as subdomains (comprising motifs of the type $CX_2CX_{28/30}HX_2C$), which are involved in the Zn^{2+}-dependent recognition of strand breaks. The C-terminal catalytic domain comprises a block of about 50 amino acids (residue 859-908), which is 100% conserved among vertebrates. This block binds the natural substrate NAD^+ and thus governs the synthesis of poly(ADP-

ribose) (cf. de Murcia, loc.cit.). The GX_3GKG motif in particular is characteristic of PARP in this block.

The beneficial function described above contrasts with a pathological one in numerous diseases (stroke, myocardial infarct, sepsis etc.). PARP is involved in cell deaths resulting from ischaemia of the brain (Choi, D.W., (1997) Nature medicine, 3, 10, 1073), of the myocardium (Zingarelli, B., et al (1997), Cardiovascular Research, 36, 205) and of the eye (Lam, T.T. (1997), Res. Comm. in Molecular Pathology and Pharmacology, 95, 3, 241). PARP activation induced by inflammatory mediators has also been observed in septic shock (Szabo, C., et al. (1997), Journal of Clinical Investigation, 100, 3, 723). In these cases, activation of PARP is accompanied by extensive consumption of NAD^+. Since four moles of ATP are consumed for the biosynthesis of one mole of NAD^+, the cellular energy supply decreases drasticallly. The consequence is cell death.

PARP inhibitors described in the abovementioned specialist literature are nicotinamide and 3-aminobenzamide. 3,4-Dihydro-5[4-1(1-piperidinyl)butoxy]-1(2H)-isoquinolone [sic] is disclosed by Takahashi, K., et al (1997), Journal of Cerebral Blood Flow and Metabolism 17, 1137. Further inhibitors are described, for example, in Banasik, M., et al. (1992) J. Biol. Chem., 267, 3, 1569 and Griffin, R.J., et al. (1995), Anti-Cancer Drug Design, 10, 507.

High molecular weight binding partners described for human PARP include the base excision repair (BER) protein XRCC1 (X-ray repair cross-complementing 1) which binds via a zinc finger motif and a BRCT (BRCA1 C terminus) module (amino acids 372-524) (Masson, M., et al., (1998) Molecular and Cellular Biology, 18,6, 3563).

It is an object of the present invention, because of the diverse physiological and pathological functions of PARP, to provide novel PARP homologs. The reason for this is that the provision of homologous PARPs would be particularly important for developing novel targets for drugs, and novel drugs, in order to improve diagnosis and/or therapy of pathological states in which PARP, PARP homologs or substances derived therefrom are involved.

We have found that this object is achieved by providing PARP homologs having an amino acid sequence which comprises

a) a functional NAD^+ binding domain
and

b) especially in the N-terminal sequence region, i.e. in the region of the first 200, such as, for example, in the region of the first 100, N-terminal amino acids, no PARP zinc finger sequence motifs of the general formula

$$CX_2CX_mHX_2C$$

in which

m is an integral value from 28 or 30, and the X radicals are, independently of one another, any amino acid;

and the functional equivalents thereof.

Thus an essential characteristic of the PARPs according to the invention is the presence of a functional NAD^+ binding domain which is located in the C-terminal region of the amino acid sequence (i.e. approximately in the region of the last 400, such as, for example, the last 350 or 300) C-terminal amino acids), in combination with an N-terminal sequence having no zinc finger motifs. Since the zinc finger motifs in known PARPs presumably contribute to recognition of the DNA breakages, it is to be assumed that the proteins according to the invention interact with DNA in another way, if at all.

The functional NAD^+ binding domain (i.e. catalytic domaine) binds the substrate for poly-ADP-ribose synthesis. Consistent with known PARPs, the sequence motif $GX_1X_2X_3GKG$, in which G is glycine, K is lysine, and X_1, X_2 and X_3 are, independently of one another, any amino acid, is present in particular. However, as shown, surprisingly, by comparison of the amino acid sequences of the NAD^+ binding domains of PARP molecules according to the invention with previously disclosed human PARP (referred to as "human PARP1" hereinafter), the sequences according to the invention differ markedly from the known sequence for the NAD^+ binding domain.

A group of PARP molecules which is preferred according to the invention preferably has the following general sequence motif in the catalytic domain in common:

$$PX_n(S/T)GX_3GKGIYFA, \text{ in particular}$$
$$(S/T)XGLRIXPX_n(S/T)GX_3GKGIYFA, \text{ preferably}$$
$$LLWHG(S/T)X_7IL(S/T)XGLRIXPX_n(S/T)GX_3GKGIYFAX_3SKSAXY$$

in which (S/T) describes the alternative occupation of this sequence position by S or T, and n is an integral value from 1 to 5, and the X radicals are, independently of one another, any amino acid. The last motif is also referred to as the "PARP signature" motif.

The automodification domain is preferably likewise present in the PARPs according to the invention. It can be located, for example, in the region from about 100 to 200 amino acids in front of the N-terminal end of the NAD^+ binding domain.

A group of preferred PARP homologs according to the invention has the additional characteristic that it comprises, N-terminally of the NAD^+ binding domain (i.e. about 30 to about 80 amino acids closer to the N terminus), a leucine zipper-like sequence motif of the general formula

$$(L/V)X_6LX_6LX_6L$$

in which
(L/V) represents the alternative occupation of this sequence position by L or V, and the X radicals are, independently of one another, any amino acid. The leucine zipper motifs observed according to the invention differ distinctly in position from those described for PARP from Drosophila. Leucine zippers may lead to homodimers (two PARP molecules) or heterodimers (one PARP molecule with a binding partner differing therefrom).

The PARP homologs according to the invention preferably additionally comprise, N-teminally of the abovementioned leucine zipper-like sequence motifs, i.e. about 10 to 250 amino acid residues closer to the N teminus, at least another one of the following part-sequence motifs:

$LX_9NX_2YX_2QLLXDX_bWGRVG$,	(motif 1)
$AX_3FXKX_4KTXNXWX_5FX_3PXK$,	(motif 2)
$QXLIX_2IX_9MX_{10}PLGKLX_3QIX_6L$,	(motif 3)
$FYTXIPHXFGX_3PP$,	(motif 4) and
$KX_3LX_2LXDIEXAX_2L$	(motif 5),

in which b is the integral value 10 or 11, and the X radicals are, independently of one another, any amino acid. It is most preferred for these motifs 1 to 5 all to be present in the stated sequence, with motif 1 being closest to the N terminus.

The abovementioned PARP signature motif is followed in the proteins according to the invention by at least another one of the following motifs:

$GX_3LXEVALG$	(motif 6)
$GX_2SX_4GX_3PX_aLXGX_2V$	(motif 7) and
$EYX_2YX_3QX_4YLL$	(motif 8)

in which a is equal to 7 to 9 and X is in each case any amino acid. It is most preferred for the three C-terminal motifs all to be present and in the stated sequence, with motif 8 being closest to the C terminus.

A preferred PARP structure according to the invention may be described schematically as follows:

Motifs 1 to 5/leucine zipper/PARP signature/motifs 6 to 8

it being possible for further amino acid residues, such as, for example, up to 40, to be arranged between the individual motifs and for further amino acid residues, such as, for example, up to 80, to be arranged at the N terminus and/or at the C terminus.

PARP homologs which are particularly preferred according to the invention are the proteins humanPARP2 and humanPARP3 and the functional equivalents thereof. The proteins [sic] referred to as humanPARP2 comprises 570 amino acids (cf. SEQ ID NO:2). The protein referred to as humanPARP3 possibly exists in two forms. Type 1 comprises 533 amino acids (SEQ ID NO:4) and Type 2 comprises 540 amino acids (SEQ ID NO:6).

The invention further relates to the binding partners for the PARP homologs according to the invention. These binding partners are preferably selected from

a) antibodies and fragments such as, for example, Fv, Fab, $(Fab)'_2$ [sic], thereof
b) protein-like compounds which interact, for example via the above leucine zipper region or another sequence section, with PARP, and
c) lower molecular weight effectors which modulate a biological PARP function such as, for example, the catalytic PARP activity, i.e. NAD^+-consuming ADP ribosylation, or the binding to an activator protein or to DNA.

The invention further relates to nucleic acids comprising

a) a nucleotide sequence coding for at least one PARP homolog according to the invention, or the complementary nucleotide sequence thereof;
b) a nucleotide sequence which hybridizes with a sequence as specified in a), preferably under stringent conditions; or
c) nucleotide sequences which are derived from the nucleotide sequences defined in a) and b) through the degeneracy of the genetic code.

Nucleic acids which are suitable according to the invention comprise in particular at least one of the part-sequences which code for the abovementioned amino acid sequence motifs.

Nucleic acids which are preferred according to the invention comprise nucleotide sequences as shown in SEQ ID NO: 1 and 3, and, in particular, part-sequences thereof which are characteristic of PARP homologs according to the invention, such as, for example, nucleotide sequences comprising

a) nucleotides +3 to +1715 shown in SEQ ID NO:1;
b) nucleotides +242 to +1843 shown in SEQ ID NO:3; or
c) nucleotides +221 to +1843 shown in SEQ ID NO:5;

or part-sequences of a), b) and c) which code for the abovementioned characteristic amino acid sequence motifs of the PARP homologs according to the invention.

The invention further relates to expression cassettes which comprise at least one of the above-described nucleotide sequences according to the invention under the genetic control of regulatory nucleotide sequences. These can be used to prepare recombinant vectors according to the invention, such as, for example, viral vectors or plasmids, which comprise at least one expression cassette according to the invention.

Recombinant microorganisms according to the invention are transformed with at least one of the abovementioned vectors.

The invention also relates to transgenic mammals transfected with a vector according to the invention.

Also provided according to the invention is an in vitro screening method for binding partners for PARP, in particular for a PARP homolog according to the invention. A first variant is carried out by

a1) immobilizing at least one PARP homolog on a support;
b1) contacting the immobilized PARP homolog with an analyte in which at least one binding partner is suspected; and
c1) determining, where appropriate after an incubation period, analyte constituents bound to the immobilized PARP homolog.

A second variant entails

a2) immobilizing an analyte which comprises at least one possible binding partner for the PARP homolog on a support;
b2) contacting the immobilized analyte with at least one PARP homolog for which a binding partner is sought; and

c3) examining the immobilized analyte, where appropriate after an incubation period, for binding of the PARP homolog.

The invention also relates to a method for the qualitative or quantitative determination of a PARP homolog-encoding nucleic acid, which comprises

a) incubating a biological sample with a defined amount of an exogenous nucleic acid according to the invention (e.g. with a length of about 20 to 500 bases or longer), hybridizing, preferably under stringent conditions, determining the hybridizing nucleic acids and, where appropriate, comparing with a standard; or

b) incubating a biological sample with a defined amount of oligonucleotide primer pairs with specificity for a PARP homolog-encoding nucleic acid, amplifying the nucleic acid, determining the amplification product and, where appropriate, comparing with a standard.

The invention further relates to a method for the qualitative or quantitative determination of a PARP homolog according to the invention, which comprises

a) incubating a biological sample with at least one binding partner specific for a PARP homolog,

b) detecting the binding partner/PARP complex and, where appropriate,

c) comparing the result with a standard.

The binding partner in this case is preferably an anti-PARP antibody or a binding fragment thereof, which carries a detectable label where appropriate.

The determination methods according to the invention for PARP, in particular for PARP homologs and for the coding nucleic acid sequences thereof, are suitable and advantageous for diagnozing sepsis- or ischemia-related tissue damage, in particular strokes, myocardial infarcts or septic shock.

The invention further comprises a method for determining the efficacy of PARP effectors, which comprises

a) incubating a PARP homolog according to the invention with an analyte which comprises an effector of a physiological or pathological PARP activity; removing the effector again where appropriate; and

b) determining the activity of the PARP homolog, where appropriate after adding substrates or cosubstrates.

The invention further relates to gene therapy compositions which comprise, in a vehicle acceptable for gene therapy, a nucleic acid construct which

a) comprises an antisense nucleic acid against a coding nucleic acid according to the invention ; or
b) a ribozyme against a noncoding nucleic acid according to the invention; or
c) codes for a specific PARP inhibitor.

The invention further relates to pharmaceutical compositions comprising, in a pharmaceutically acceptable vehicle, at least one PARP protein according to the invention, at least one PARP binding partner according to the invention or at least one coding nucleotide sequence according to the invention.

Finally, the invention relates to the use of low molecular weight (less than about 1000 Dalton) binding partners of a PARP homolog for diagnosis or therapy of pathological states in the development and/or progress of which at least one one PARP protein, in particular a PARP homolog according to the invention, or a polypeptide derived therefrom, are [sic] involved.

The present invention will now be described in more detial with reference to the appended figures. These show:

In Figure 1 a sequence alignment of human PARP (humanPARP1) and two PARPs preferred according to the invention (humanPARP2 and humanPARP3). Sequence agreements between humanPARP1 and humanPARP2 or humanPARP3 are depicted within frames. The majority sequence is indicated over the alignment. The zinc finger motifs of humanPARP1 are located in the sequence sections corresponding to amnio [sic] acid residues 21 to 56 and 125 to 162;

In Figure 2 a Northern blot with various human tissues to illustrate the tissue distribution of PARP molecules according to the invention. Lane 1: heart; lane 2: brain; lane 3: placenta; lane 4: lung; lane 5: liver; lane 6: skeletal muscle; lane 7: kidney; lane 8: pancreas; (A) blot for humanPARP2; (B) blot for humanPARP3; the respective positions of the size standards (kb) are indicated between (A) and (B).

PARP homologs and functional equivalents

Unless stated otherwise, for the purposes of the present description amino acid sequences are indicated starting with the N terminus. If the one-letter code is used for amino acids, then G is glycine, A is alanine, V is valine, L is leucine, I is

isoleucine, S is serine, T is threonine, D is aspartic acid, N is asparagine, E is glutamic acid, Q is glutamine, W is tryptophan, H is histidine, R is arginine, P is proline, K is lysine, Y is tyrosine, F is phenylalanine, C is cysteine and M is methionine.

The present invention is not confined to the PARP homologs specifically described above. On the contrary, those homologs which are functional equivalents thereof are also embraced. Functional equivalents comprise both natural, such as, for example, species-specific or organ-specific, and artificially produced variants of the proteins specifically described herein. Functional equivalents according to the invention differ by addition, substitution, inversion, insertion and/or deletion of one or more amino acid residues of humanPARP2 (SEQ ID NO:2) and humanPARP3 (SEQ ID NO: 4 and 6), there being at least retention of the NAD-binding function of the protein mediated by a functional catalytic C-terminal domain. Functional equivalents also comprise where appropriate those variants in which the leucine zipper region is essentially retained.

It is moreover possible, for example, starting from the sequence for humanPARP2 or humanPARP3 to replace certain amino acids by those with similar physicochemical properties (bulk, basicity, hydrophobicity, etc.). It is possible, for example, for arginine residues to be replaced by lysine residues, valine residues by isoleucine residues or aspartic acid residues by glutamic acid residues. However, it is also possible for one or more amino acids to be exchanged in sequence, added or deleted, or several of these measures can be combined together. The proteins which have been modified in this way from the humanPARP2- or humanPARP3 sequence have at least 60%, preferably at least 75%, very particularly preferably at least 85%, homology with the starting sequence, calculated using the algorithm of Pearson and Lipman, Proc. Natl. Acad. Sci (USA) 85(8), 1988, 2444-2448.

The following homologies have been determined at the amino acid level and DNA level between humanPARP1, 2 and 3 (FastA program, Pearson and Lipman, loc. cit.):
Amino acid homologies:

	Percent identity	Percent identity in PARP signature
PARP1/PARP2	41.97% (517)	86% (50)
PARP1/PARP3	33.81% (565)	53.1% (49)
PARP2/PARP3	35.20% (537)	53.1% (49)

Numbers in parentheses indicate the number of overlapping amino acids.

DNA Homologies:

	Percent identity in the ORF	Percent identity in PARP signature
PARP1/PARP2	60.81% (467)	77.85% (149)
PARP1/PARP3	58.81% (420)	59.02% (61)
PARP2/PARP3	60.22% (269)	86.36% (22)

Numbers in parentheses indicate the number of overlapping nucleotides.

The polypeptides according to the invention can be classified as homologous poly(ADP-ribose) polymerases on the basis of the great similarity in the region of the catalytic domain.

It is also essential to the invention that the novel PARP homologs do not have conventional zinc finger motifs. This means that these enzymes is [sic] not necessarily involved in DNA repair, but are still able to carry out their pathological mechanism (NAD^+ consumption and thus energy consumption due to ATP consumption). This is particularly important for drug development. Potential novel inhibitors of the polymerases according to the invention can thus inhibit the pathological functions without having adverse effects on the desired physiological properties. This was impossible with inhibitors against the PARPs known to date since there was always also inhibition of the DNA repair function. The potentially mutagenic effect of known PARP inhibitors is thus easy to understand.

The PARP homolog which is preferred according to the invention and is shown in SEQ ID NO:2 (human PARP2) can advantageously be

isolated from human brain. The expression of human PARP2 in other tissues or organs is distinctly weaker.

The PARP homolog which is preferred according to the invention and is shown in SEQ ID NO: 4 and 6 (humanPARP3) can advantageously be isolated from human brain, heart or kidney. The expression of humanPARP3 in other tissues or organs, such as muscle or liver, is distinctly weaker.

The skilled worker familiar with protein isolation will make use of the combination of preparative methodologies which is most suitable in each case for isolating natural PARPs according to the invention from tissues or recombinantly prepared PARPs according to the invention from cell cultures. Suitable standard preparative methods are described, for example, in Cooper, T.G., Biochemische Arbeitsmethoden, published by Walter de Gruyter, Berlin, New York or in Scopes, R. Protein Purification, Springer Verlag, New York, Heidelberg, Berlin.

The invention additionally relates to PARP2 and PARP3 homologs which, although they can be isolated from other eukaryotic species, i.e. invertebrates or vertebrates, especially other mammals such as, for example, mice, rats, cats, dogs, pigs, sheep, cattle, horses or monkeys, or from other organs such as, for example the myocardium, have the essential structural and functional properties predetermined by the PARPs according to the invention.

In particular, the humanPARP2 which can be isolated from human brain, and its functional equivalents, are preferred agents for developing inhibitors of stroke. This is because it can be assumed that drug development based on PARP2 as indicator makes it possible to develop inhibitors which are optimized for use in the human brain. However, it cannot be ruled out that inhibitors developed on the basis of PARP2 can also be employed for treating PARP-mediated pathological states in other organs too.

Another essential biological property of PARPs according to the invention and their functional equivalents is to be seen in their ability to bind an interacting partner. HumanPARP2 and 3 differ from previously disclosed PARPs from higher eukaryotes such as, in particular, mammals by having so-called leucine zipper motifs. This is a typical motif for protein-protein interactions. It is possible that these motifs permit PARP activation by an interacting partner. This additional structural element thus also

provides a possible starting point for development of PARP effectors such as, for example, inhibitors.

The invention thus further relates to proteins which interact with PARP2 and/or 3, preferably those which bring about their activation or inactivation.

The invention further relates to proteins which still have the abovementioned ligand-binding activity and which can be prepared starting from the specifically disclosed amino acid sequences by targeted modifications.

It is possible, starting from the peptide sequence of the proteins according to the invention, to generate synthetic peptides which are employed, singly or in combination, as antigens for producing polyclonal or monoclonal antibodies. It is also possible to employ the PARP protein or fragments thereof for generating antibodies. The invention thus also relates to peptide fragments of PARP proteins according to the invention which comprise characteristic part-sequences, in particular those oligo- or polypeptides which comprises [sic] at least one of the abovementioned sequence motifs. Fragments of this type can be obtained, for example, by proteolytic digestion of PARP proteins or by chemical synthesis of peptides.

Nucleic acids coding for PARP homologs:

Unless stated otherwise, nucleotide sequences are indicated in the present description from the 5' to the 3' direction.

The invention further relates to nucleic acid sequences which code for the abovementioned proteins, in particular for those having the amino acid sequence depicted in SEQ ID NO: 2, 4 and 6, but without being restricted thereto. Nucleic acid sequences which can be used according to the invention also comprise allelic variants which, as described above for the amino acid sequences, are obtainable by deletion, inversion, insertion, addition and/or substitution of nucleotides, preferably of nucleotides shown in SEQ ID NO: 1 and 3, but with essential retention of the biological properties and the biological activity of the corresponding gene product. Nucleotide sequences which can be used are obtained, for example, by silent (without alteration of the amino acid sequence) or conservative (exchange of amino acids of the same size, charge, polarity or solubility) nucleotide substitutions.

Nucleic acid sequences according to the invention also embrace functional equivalents of the genes, such as eukaryotic homologs for example from invertebrates such as Caenorhabditis or Drosophila, or vertebrates, preferably from the mammals described above. Preferred genes are those from vertebrates which code for a gene product which has the properties essential to the invention as described above.

The nucleic acids according to the invention can be obtained in a conventional way by various routes:

For example, a genomic or a cDNA library can be screened for DNA which codes for a PARP molecule or a part thereof. For example, a DNA library obtained from human brain, heart or kidney can be screened with a suitable probe such as, for example, a labeled single-stranded DNA fragment which corresponds to a part-sequence of suitable length selected from SEQ ID NO: 1 or 3, or sequence complementary thereto. For this purpose, it is possible, for example, for the DNA fragments of the library which have been transferred into a suitable cloning vector to be, after transformation into a bacterium, plated out on agar plates. The clones can then be transferred to nitrocellulose filters and, after denaturation of the DNA, hybridized with the labeled probe. Positive clones are then isolated and characterized.

The DNA coding for PARP homologs according to the invention or partial fragments can also be synthesized chemically starting from the sequence information contained in the present application. For example, it is possible for this purpose for oligonucleotides with a length of about 100 bases to be synthesized and sequentially ligated in a manner known per se by, for example, providing suitable terminal restriction cleavage sites.

The nucleotide sequences according to the invention can also be prepared with the aid of the polymerase chain reaction (PCR). For this, a target DNA such as, for example, DNA from a suitable full-length clone is hybridized with a pair of synthetic oligonucleotide primers which have a length of about 15 bases and which bind to opposite ends of the target DNA. The sequence section lying between them is then filled in with DNA polymerase. Repetition of this cycle many times allows the target DNA to be amplified (cf. White et al.(1989), Trends Genet. 5, 185).

The nucleic acid sequences according to the invention are also to be understood to include truncated sequences, single-stranded DNA or RNA of the coding and noncoding, complementary DNA sequence, mRNA sequences and cDNAs derived therefrom.

The invention further embraces nucleotide sequences hybridizing with the above sequences under stringent conditions. Stringent hybridization conditions for the purpose of the present invention exist when the hybridizing sequences have a homology of about 70 to 100%, such as, for example about 80 to 100% or 90 to 100% (preferably in an amino acid section of at least about 40, such as, for example, about 50, 100, 150, 200, 400 or 500 amino acids).

Stringent conditions for the screening of DNA, in particular cDNA banks, exist, for example, when the hybridization mixture is washed with 0.1X SSC buffer (20X SSC buffer = 3M NaCl, 0.3M sodium citrate, pH 7.0) and 0.1% SDS at a temperature of about 60°C.

Northern blot analyses are analyses are washed under stringent conditions with 0.1X SSC, 0,1% SDS at a temperature of about 68°C, for example.

Nucleic acid derivatives and expression constructs:

The nucleic acid sequences are also to be understood to include derivatives such as, for example, promoter variants or alternative splicing variants. The promoters operatively linked in front of the nucleotide sequences according to the invention may moreover be modified by nucleotide addition(s) or substitution(s), inversion(s), insertion(s) and/or deletion(s), but without impairing the functionality or activity of the promoters. The promoters can also have their activity increased by modifying their sequence, or be completely replaced by more effective promoters even from heterologous organisms. The promoter variants described above are used to prepare expression cassettes according to the invention.

Specific examples of human PARP2 splicing variants which may be mentioned are:

Variant humanPARP2a: Deletion of base pairs 766 to 904 (cf. SEQ ID NO:1). This leads to a frame shift with a new stop codon ("TAA" corresonding to nucleotides 922 to 924 in SEQ ID NO:1).
Variant humanPARP2b: Insertion of
5'- gta tgc cag gaa ggt cat ggg cca gca aaa ggg tct ctg -3'

after nucleotide 204 (SEQ ID NO:1). This extends the amino acid sequence by the insertion: GMPGRSWASKRVS

Nucleic acid derivatives also mean variants whose nucleotide sequence [sic] in the region from -1 to -1000 in front of the start codon have been modified so that gene expression and/or protein expression is increased.

Besides the nucleotide sequence described above, the nucleic acid constructs which can be used according to the invention comprise in functional, operative linkage one or more other regulatory sequences, such as promoters, amplification signals, enhancers, polyadenylation sequences, origins of replication, reporter genes, selectable marker genes and the like. This linkage may, depending on the desired use, lead to an increase or decrease in gene expression.

In addition to the novel regulatory sequences, it is possible for the natural regulatory sequence still to be present in front of the actual structural genes. This natural regulation can, where appropriate, be switched off by genetic modification, and the expression of the genes increased or decreased. However, the gene construct may also have a simpler structure, that is to say no additional regulatory signals are inserted in front of the structural genes, and the natural promoter with its regulation is not deleted. Instead, the natural regulatory sequence is mutated in such a way that regulation no longer takes place, and gene expression is enhanced or diminished. It is also possible to insert additional advantageous regulatory elements at the 3' end of the nucleic acid sequence. The nucleic acid sequences can be present in one or more copies in the gene construct.

Advantageous regulatory sequences for the expression method according to the invention are, for example, present in promoters such as cos, tac, trp, tet, trp-tet, lpp, lac, lpp-lac, lacIq, T7, T5, T3, gal, trc, ara, SP6, l-PR or the l-PL promoter, which are advantageously used in Gram-negative bacteria. Other advantageous regulatory sequences are present, for example, in the Gram-positive promoters amy and SPO2, in the yeast promoters ADC1, MFa , AC, P-60, CYC1, GAPDH or in the plant promoters CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or in the ubiquitin or phaseolin promoter.

It is possible in principle to use all natural promoters with their regulatory sequences. It is also possible and advantageous to use synthetic promoters.

Said regulatory sequences are intended to make specific expression of the nucleic acid sequences and of [sic] protein expression possible. This may mean, for example, depending on the host organism that the gene is expressed or overexpressed only after induction, or that it is immediately expressed and/or overexpressed.

The regulatory sequences or factors may moreover preferably have a positive influence on, and thus increase or decrease, the expression. Thus, enhancement of the regulatory elements may advantageously take place at the level of transcription by using strong transcription signals such as promoters and/or enhancers. However, it is also possible to enhance translation by, for example, improving the stability of the mRNA.

Enhancers mean, for example, DNA seqences which bring about increased expression via an improved interaction between RNA polymerase and DNA.

The recombinant nucleic acid construct or gene construct is, for expression in a suitable host organism, advantageously inserted into a host-specific vector which makes optimal expression of the genes in the host possible. Vectors are well known to the skilled worker and are to be found, for example, in "Cloning Vectors" (Pouwels P. H. et al., Ed., Elsevier, Amsterdam-New York-Oxford, 1985). Apart from plasmids, vectors also mean all other vectors known to the skilled worker, such as, for example, phages, viruses, such as SV40, CMV, baculovirus and adenovirus, transposons, IS elements, phasmids, cosmids, and linear or circular DNA. These vectors may undergo autonomous replication in the host organism or chromosomal replication.

Expression of the constructs:

The recombinant constructs according to the invention described above are advantageously introduced into a suitable host system and are expressed. Cloning and transfection methods familiar to the skilled worker are preferably used in order to bring about expression of said nucleic acids in the particular expression system. Suitable systems are described, for example, in Current Protocols in Molecular Biology, F. Ausubel et al., ed., Wiley Interscience, New York 1997.

Suitable host organisms are in principle all organisms which make it possible to express the nucleic acids according to the invention, their allelic variants, their functional equivalents or derivatives or the recombinant nucleic acid construct. Host

organisms mean, for example, bacteria, fungi, yeasts, plant or animal cells. Preferred organisms are bacteria such as those of the genera Escherichia, such as, for example, Escherichia coli, Streptomyces, Bacillus or Pseudomonas, eukaryotic microorganisms such as Saccharomyces cerevisiae, Aspergillus, higher eukaryotic cells from animals or plants, for example Sf9 or CHO cells.

The gene product can also, if required, be expressed in transgenic organisms such as transgenic animals such as, in particular, mice, sheep, or transgenic plants. The transgenic organisms may also be so-called knock-out animals or plants in which the corresponding endogenous gene has been switched off, such as, for example, by mutation or partial or complete deletion.

The combination of the host organisms and the vectors appropriate for the organisms, such as plasmids, viruses or phages, such as, for example, plasmids with the RNA polymerase/promoter system, phages λ, μ or other temperate phages or transposons and/or other advantageous regulatory sequences form [sic] an expression system. The term expression systems preferably means, for example, a combination of mammalian cells such as CHO cells, and vectors, such as pcDNA3neo vector, which are suitable for mammalian cells.

As described above, the gene product can also be expressed advantageously in transgenic animals, e.g. mice, sheep, or transgenic plants. It is likewise possible to program cell-free translation systems with the RNA derived from the nucleic acid.

The gene product can also be expressed in the form of therapeutically or diagnostically suitable fragments. To isolate the recombinant protein it is possible and advantageous to use vector systems or oligonucleotides which extend the cDNA by certain nucleotide sequences and thus code for modified polypeptides which serve to simplify purification. Suitable modifications of this type are, for example, so-called tags which act as anchors, such as, for example, the modification known as the hexa-histidine anchor, or epitopes which can be recognized as antigens by antibodies (described, for example, in Harlow, E. and Lane, D., 1988, Antibodies: A Laboratory Manual. Cold Spring Harbor (N.Y.) Press). These anchors can be used to attach the proteins to a solid support such as, for example, a polymer matrix, which can, for example, be packed into a chromatography column, or to a microtiter plate or to another support.

These anchors can also at the same time be used to recognize the proteins. It is also possible to use for recognition of the proteins conventional markers such as fluorescent dyes, enzyme markers which form a detectable reaction product after reaction with a substrate, or radioactive markers, alone or in combination with the anchors for derivatizing the proteins.

Production of antibodies:

Anti-PARP2 antibodies are produced in a manner familiar to the skilled worker. Antibodies mean both polyclonal, monoclonal, human or humanized antibodies or fragments thereof, single chain antibodies or else synthetic antibodies, likewise antibody fragments such as Fv, Fab and $(Fab)'_2$ [sic]. Suitable production methods are described, for example, in Campbell, A.M., Monoclonal Antibody Technology, (1987) Elsevier Verlag, Amsterdam, New York, Oxford and in Breitling, F. and Dübel, S., Rekombinante Antikörper (1997), Spektrum Akademischer Verlag, Heidelberg.

Further use of the coding sequence:

The present cDNA additionally provides the basis for cloning the genomic sequence of the novel PARP gene. This also includes the relevant regulatory or promoter sequence, which is available, for example, by sequencing the region located 5' upstream of the cDNA according to the invention. The cDNA sequence information is also the basis for producing antisense molecules or ribozymes with the aid of known methods (cf. Jones, J.T. and Sallenger, B.A. (1997) Nat. Biotechnol. 15, 902; Nellen, W. and Lichtenstein, C. (1993) TIBS, 18, 419). Die genomische DNA can likewise be used to produce the gene constructs described above.

Another possibility of using the nucleotide sequence or parts thereof is to generate transgenic animals. Transgenic overexpression or genetic knock-out of the sequence information in suitable animal models may provide further valuable information about the (patho)physiology of the novel enzymes.

Therapeutic applications:

In situations where there is a prevailing deficiency of a protein according to the invention it is possible to employ several methods for replacement. On the one hand, the protein, natural or recombinant, can be administered directly or by gene therapy in the form of its coding nucleic acid (DNA or RNA). It is possible to use any suitable vectors for this, for example both viral and non-viral vehicles. Suitable methods are described, for example,

by Strauss and Barranger in Concepts in Gene Therapy (1997), Walter de Gruyter, publisher. Another alternative is provided by stimulation of the endogenous gene by suitable agents.

It is also possible to block the turnover or the inactivation of PARPs according to the invention, for example by proteases. Finally, inhibitors or agonists of PARPs according to the invention can be employed.

In situations where a PARP is present in excess, various types of inhibitors can be employed. This inhibition can be achieved both by antisense molecules, ribozymes, oligonucleotides or antibodies, and by low molecular weight compounds.

Nontherapeutic applications:

The nucleic acids according to the invention, such as, for example, cDNA, the genomic DNA, the promoter, and the polypeptide, and partial fragments thereof, can also be used in recombinant or nonrecombinant form for developing various test systems.

For example, it is possible to establish a test system which is suitable for measuring the activity of the promoter or of the protein in the presence of a test substance. The methods of measurement in this case are preferably simple ones, e.g. colorimetric, luminometric, fluorimetric, immunological or radioactive, and allow preferably a large number of test substances to be measured rapidly. Tests of this type are suitable and advantageous for so-called high-throughput screening. These test systems allow test substances to be assessed for their binding to or their agonism, antagonism or inhibition of proteins according to the invention.

Determination of the amount, activity and distribution of the proteins according to the invention or their underlying mRNA in the human body can be used for the diagnosis, for the determination of the predisposition and for the monitoring of certain diseases. Likewise, the sequence of the cDNA and of [sic] the genomic sequence may provide information about genetic causes of and predispositions to certain diseases. It is possible to use for this purpose both DNA/RNA probes and antibodies of a wide variety of types. The nucleotide sequences according to the invention or parts thereof can further be used in the form of suitable probes for detecting point mutations, deletions or insertions.

The proteins according to the invention can further be used to identify and isolate their natural ligands or interacting partners. The proteins according to the invention can additionally be used to identify and isolate artificial or synthetic ligands. For this purpose, the recombinantly prepared or purified natural protein can be derivatized in such a way that it has modifications which permit linkage to support materials. Proteins bound in this way can be incubated with various analytes, such as, for example, protein extracts or peptide libraries or other sources of ligands. Specifically bound peptides, proteins or low molecular weight, non-proteinogenous substances can be isolated and characterized in this way. Non-proteinogenous substances mean, for example, low molecular weight chemical substances (= less than 1000 Dalton) which may originate, for example, from classical drug synthesis or from so-called substance libraries which have been synthesized combinatorially.

The protein extracts used are derived, for example, from homogenates of plants or parts of plants, microorganisms, human or animal tissues or organs.

Ligands or interacting partners can also be identified by methods like the yeast two-hybrid system (Fields, S. and Song, O. (1989) Nature, 340, 245). The expression banks which can be employed in this case may be derived, for example, from human tissues such as, for example, brain, heart, kidney etc.

The nucleic acid sequences according to the invention and the proteins encoded by them can be employed for developing reagents, agonists and antagonists or inhibitors for the diagnosis and therapy of chronic and acute diseases associated with the expression of one of the protein sequences according to the invention, such as, for example, with increased or decreased expression thereof. The reagents, agonists, antagonists or inhibitors developed can subsequently be used to produce pharmaceutical preparations for the treatment or diagnosis of disorders. Examples of possible diseases in this connection are those of the brain, of the cardiovascular system or of the eye, or septic shock.

The invention is now illustrated in detail with reference to the following examples.

Example 1: Isolation of the PARP2- and PARP3-cDNA

The present cDNA sequences were found for the first time on sequence analysis of cDNA clones of a cDNA library from human brain (Human Brain 5'Stretch Plus cDNA Library, # HL3002a, Clontech). The sequences of these clones are described in SEQ ID NO:1 and 3.

Example 2: Expression of humanPARP2 and humanPARP3 in human tissues

The expression of humanPARP2 and humanPARP3 was investigated in eight different human tissues by northern blot analysis. A Human Multiple Tissue Northern Blot supplied by Clontech (#7760-1) was hybridized for this purpose with an RNA probe. The probe was produced by *in vitro* transcription of the corresponding cDNA of human PARP2 and human PARP3 in the presence of digoxigenin-labeled nucleotides.

After stringent washing, the transcript of human PARP2 was mainly detected in brain, but there is slight expression also in the heart. Expression in other tissues (placenta, lung, liver, skeletal muscle, kidney, pancreas) is very weak. The transcript size of about 1.9 kb corresponds to the length of the cDNA determined (1.85kb) (cf. Figure 2(A)).

After stringent washing, the transcript of human PARP3 was mainly detected in heart, brain and kidney, and it is likewise expressed distinctly, but weaker, in skeletal muscle and liver. Expression in other tissues (placenta, lung, pancreas) is distinctly weaker (cf. Figure 2(B)). There are at least 2 transcripts for humanPARP3. Their size (about 2.2 kb and 2.5 kb respectively) corresponds to the length of the cDNA determined (2.3kb).

A 0.1X SSC buffer (prepared from 20X SSC: 3M NaCl, 0.3M sodium citrate, pH 7.0) supplemented with 0.1% SDS was used for the stringent washing at 68°C.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: BASF Aktiengesellschaft

(B) STREET: -

(C) CITY: Ludwigshafen

(E) COUNTRY: Germany

(F) POSTAL CODE: 67065

(ii) TITLE OF INVENTION: Novel poly(ADP-ribose) polymerase genes

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER-READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1843 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(vi) ORIGINAL SOURCE:

(F) TISSUE TYPE: brain

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 3..1715

(D) OTHER INFORMATION:/product= "Poly ADP Ribose Polymerase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CC ATG GCG GCG CGG CGG CGA CGG AGC ACC GGC GGC GGC AGG GCG AGA 47
Met Ala Ala Arg Arg Arg Arg Ser Thr Gly Gly Gly Arg Ala Arg

```
     1               5                  10                  15

GCA TTA AAT GAA AGC AAA AGA GTT AAT AAT GGC AAC ACG GCT CCA GAA      95
Ala Leu Asn Glu Ser Lys Arg Val Asn Asn Gly Asn Thr Ala Pro Glu
                20                  25                  30

GAC TCT TCC CCT GCC AAG AAA ACT CGT AGA TGC CAG AGA CAG GAG TCG     143
Asp Ser Ser Pro Ala Lys Lys Thr Arg Arg Cys Gln Arg Gln Glu Ser
            35                  40                  45

AAA AAG ATG CCT GTG GCT GGA GGA AAA GCT AAT AAG GAC AGG ACA GAA     191
Lys Lys Met Pro Val Ala Gly Gly Lys Ala Asn Lys Asp Arg Thr Glu
        50                  55                  60

GAC AAG CAA GAT GAA TCT GTG AAG GCC TTG CTG TTA AAG GGC AAA GCT     239
Asp Lys Gln Asp Glu Ser Val Lys Ala Leu Leu Leu Lys Gly Lys Ala
    65                  70                  75

CCT GTG GAC CCA GAG TGT ACA GCC AAG GTG GGG AAG GCT CAT GTG TAT     287
Pro Val Asp Pro Glu Cys Thr Ala Lys Val Gly Lys Ala His Val Tyr
80                  85                  90                  95

TGT GAA GGA AAT GAT GTC TAT GAT GTC ATG CTA AAT CAG ACC AAT CTC     335
Cys Glu Gly Asn Asp Val Tyr Asp Val Met Leu Asn Gln Thr Asn Leu
                100                 105                 110

CAG TTC AAC AAC AAC AAG TAC TAT CTG ATT CAG CTA TTA GAA GAT GAT     383
Gln Phe Asn Asn Asn Lys Tyr Tyr Leu Ile Gln Leu Leu Glu Asp Asp
            115                 120                 125

GCC CAG AGG AAC TTC AGT GTT TGG ATG AGA TGG GGC CGA GTT GGG AAA     431
Ala Gln Arg Asn Phe Ser Val Trp Met Arg Trp Gly Arg Val Gly Lys
        130                 135                 140

ATG GGA CAG CAC AGC CTG GTG GCT TGT TCA GGC AAT CTC AAC AAG GCC     479
Met Gly Gln His Ser Leu Val Ala Cys Ser Gly Asn Leu Asn Lys Ala
    145                 150                 155

AAG GAA ATC TTT CAG AAG AAA TTC CTT GAC AAA ACG AAA AAC AAT TGG     527
Lys Glu Ile Phe Gln Lys Lys Phe Leu Asp Lys Thr Lys Asn Asn Trp
160                 165                 170                 175

GAA GAT CGA GAA AAG TTT GAG AAG GTG CCT GGA AAA TAT GAT ATG CTA     575
Glu Asp Arg Glu Lys Phe Glu Lys Val Pro Gly Lys Tyr Asp Met Leu
                180                 185                 190

CAG ATG GAC TAT GCC ACC AAT ACT CAG GAT GAA GAG GAA ACA AAG AAA     623
```

```
Gln Met Asp Tyr Ala Thr Asn Thr Gln Asp Glu Glu Glu Thr Lys Lys
               195                 200                 205

GAG GAA TCT CTT AAA TCT CCC TTG AAG CCA GAG TCA CAG CTA GAT CTT      671
Glu Glu Ser Leu Lys Ser Pro Leu Lys Pro Glu Ser Gln Leu Asp Leu
        210                 215                 220

CGG GTA CAG GAG TTA ATA AAG TTG ATC TGT AAT GTT CAG GCC ATG GAA      719
Arg Val Gln Glu Leu Ile Lys Leu Ile Cys Asn Val Gln Ala Met Glu
    225                 230                 235

GAA ATG ATG ATG GAA ATG AAG TAT AAT ACC AAG AAA GCC CCA CTT GGG      767
Glu Met Met Met Glu Met Lys Tyr Asn Thr Lys Lys Ala Pro Leu Gly
240                 245                 250                 255

AAG CTG ACA GTG GCA CAA ATC AAG GCA GGT TAC CAG TCT CTT AAG AAG      815
Lys Leu Thr Val Ala Gln Ile Lys Ala Gly Tyr Gln Ser Leu Lys Lys
                260                 265                 270

ATT GAG GAT TGT ATT CGG GCT GGC CAG CAT GGA CGA GCT CTC ATG GAA      863
Ile Glu Asp Cys Ile Arg Ala Gly Gln His Gly Arg Ala Leu Met Glu
            275                 280                 285

GCA TGC AAT GAA TTC TAC ACC AGG ATT CCG CAT GAC TTT GGA CTC CGT      911
Ala Cys Asn Glu Phe Tyr Thr Arg Ile Pro His Asp Phe Gly Leu Arg
        290                 295                 300

ACT CCT CCA CTA ATC CGG ACA CAG AAG GAA CTG TCA GAA AAA ATA CAA      959
Thr Pro Pro Leu Ile Arg Thr Gln Lys Glu Leu Ser Glu Lys Ile Gln
    305                 310                 315

TTA CTA GAG GCT TTG GGA GAC ATT GAA ATT GCT ATT AAG CTG GTG AAA     1007
Leu Leu Glu Ala Leu Gly Asp Ile Glu Ile Ala Ile Lys Leu Val Lys
320                 325                 330                 335

ACA GAG CTA CAA AGC CCA GAA CAC CCA TTG GAC CAA CAC TAT AGA AAC     1055
Thr Glu Leu Gln Ser Pro Glu His Pro Leu Asp Gln His Tyr Arg Asn
                340                 345                 350

CTA CAT TGT GCC TTG CGC CCC CTT GAC CAT GAA AGT TAC GAG TTC AAA     1103
Leu His Cys Ala Leu Arg Pro Leu Asp His Glu Ser Tyr Glu Phe Lys
            355                 360                 365

GTG ATT TCC CAG TAC CTA CAA TCT ACC CAT GCT CCC ACA CAC AGC GAC     1151
Val Ile Ser Gln Tyr Leu Gln Ser Thr His Ala Pro Thr His Ser Asp
        370                 375                 380
```

```
TAT ACC ATG ACC TTG CTG GAT TTG TTT GAA GTG GAG AAG GAT GGT GAG      1199
Tyr Thr Met Thr Leu Leu Asp Leu Phe Glu Val Glu Lys Asp Gly Glu
    385                 390                 395

AAA GAA GCC TTC AGA GAG GAC CTT CAT AAC AGG ATG CTT CTA TGG CAT      1247
Lys Glu Ala Phe Arg Glu Asp Leu His Asn Arg Met Leu Leu Trp His
400                 405                 410                 415

GGT TCC AGG ATG AGT AAC TGG GTG GGA ATC TTG AGC CAT GGG CTT CGA      1295
Gly Ser Arg Met Ser Asn Trp Val Gly Ile Leu Ser His Gly Leu Arg
                420                 425                 430

ATT GCC CCA CCT GAA GCT CCC ATC ACA GGT TAC ATG TTT GGG AAA GGA      1343
Ile Ala Pro Pro Glu Ala Pro Ile Thr Gly Tyr Met Phe Gly Lys Gly
            435                 440                 445

ATC TAC TTT GCT GAC ATG TCT TCC AAG AGT GCC AAT TAC TGC TTT GCC      1391
Ile Tyr Phe Ala Asp Met Ser Ser Lys Ser Ala Asn Tyr Cys Phe Ala
        450                 455                 460

TCT CGC CTA AAG AAT ACA GGA CTG CTG CTC TTA TCA GAG GTA GCT CTA      1439
Ser Arg Leu Lys Asn Thr Gly Leu Leu Leu Leu Ser Glu Val Ala Leu
    465                 470                 475

GGT CAG TGT AAT GAA CTA CTA GAG GCC AAT CCT AAG GCC GAA GGA TTG      1487
Gly Gln Cys Asn Glu Leu Leu Glu Ala Asn Pro Lys Ala Glu Gly Leu
480                 485                 490                 495

CTT CAA GGT AAA CAT AGC ACC AAG GGG CTG GGC AAG ATG GCT CCC AGT      1535
Leu Gln Gly Lys His Ser Thr Lys Gly Leu Gly Lys Met Ala Pro Ser
                500                 505                 510

TCT GCC CAC TTC GTC ACC CTG AAT GGG AGT ACA GTG CCA TTA GGA CCA      1583
Ser Ala His Phe Val Thr Leu Asn Gly Ser Thr Val Pro Leu Gly Pro
            515                 520                 525

GCA AGT GAC ACA GGA ATT CTG AAT CCA GAT GGT TAT ACC CTC AAC TAC      1631
Ala Ser Asp Thr Gly Ile Leu Asn Pro Asp Gly Tyr Thr Leu Asn Tyr
        530                 535                 540

AAT GAA TAT ATT GTA TAT AAC CCC AAC CAG GTC CGT ATG CGG TAC CTT      1679
Asn Glu Tyr Ile Val Tyr Asn Pro Asn Gln Val Arg Met Arg Tyr Leu
    545                 550                 555

TTA AAG GTT CAG TTT AAT TTC CTT CAG CTG TGG TGA ATGTTGATAT           1725
Leu Lys Val Gln Phe Asn Phe Leu Gln Leu Trp  *
```

560 565 570

TAAATAAACC AGAGATCTGA TCTTCAAGCA AGAAAATAAG CAGTGTTGTA CTTGTGAATT 1785

TTGTGATATT TTATGTAATA AAAACTGTAC AGGTCTAAAA AAAAAAAAAA AAAAAAAA 1843

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 571 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Ala Arg Arg Arg Arg Ser Thr Gly Gly Gly Arg Ala Arg Ala
1 5 10 15

Leu Asn Glu Ser Lys Arg Val Asn Asn Gly Asn Thr Ala Pro Glu Asp
20 25 30

Ser Ser Pro Ala Lys Lys Thr Arg Arg Cys Gln Arg Gln Glu Ser Lys
35 40 45

Lys Met Pro Val Ala Gly Gly Lys Ala Asn Lys Asp Arg Thr Glu Asp
50 55 60

Lys Gln Asp Glu Ser Val Lys Ala Leu Leu Leu Lys Gly Lys Ala Pro
65 70 75 80

Val Asp Pro Glu Cys Thr Ala Lys Val Gly Lys Ala His Val Tyr Cys
85 90 95

Glu Gly Asn Asp Val Tyr Asp Val Met Leu Asn Gln Thr Asn Leu Gln
100 105 110

Phe Asn Asn Asn Lys Tyr Tyr Leu Ile Gln Leu Leu Glu Asp Asp Ala
115 120 125

Gln Arg Asn Phe Ser Val Trp Met Arg Trp Gly Arg Val Gly Lys Met
130 135 140

Gly Gln His Ser Leu Val Ala Cys Ser Gly Asn Leu Asn Lys Ala Lys
145 150 155 160

Glu Ile Phe Gln Lys Lys Phe Leu Asp Lys Thr Lys Asn Asn Trp Glu

```
                    165                 170                 175

Asp Arg Glu Lys Phe Glu Lys Val Pro Gly Lys Tyr Asp Met Leu Gln
                180                 185                 190

Met Asp Tyr Ala Thr Asn Thr Gln Asp Glu Glu Glu Thr Lys Lys Glu
        195                 200                 205

Glu Ser Leu Lys Ser Pro Leu Lys Pro Glu Ser Gln Leu Asp Leu Arg
    210                 215                 220

Val Gln Glu Leu Ile Lys Leu Ile Cys Asn Val Gln Ala Met Glu Glu
225                 230                 235                 240

Met Met Met Glu Met Lys Tyr Asn Thr Lys Lys Ala Pro Leu Gly Lys
                245                 250                 255

Leu Thr Val Ala Gln Ile Lys Ala Gly Tyr Gln Ser Leu Lys Lys Ile
            260                 265                 270

Glu Asp Cys Ile Arg Ala Gly Gln His Gly Arg Ala Leu Met Glu Ala
        275                 280                 285

Cys Asn Glu Phe Tyr Thr Arg Ile Pro His Asp Phe Gly Leu Arg Thr
    290                 295                 300

Pro Pro Leu Ile Arg Thr Gln Lys Glu Leu Ser Glu Lys Ile Gln Leu
305                 310                 315                 320

Leu Glu Ala Leu Gly Asp Ile Glu Ile Ala Ile Lys Leu Val Lys Thr
                325                 330                 335

Glu Leu Gln Ser Pro Glu His Pro Leu Asp Gln His Tyr Arg Asn Leu
            340                 345                 350

His Cys Ala Leu Arg Pro Leu Asp His Glu Ser Tyr Glu Phe Lys Val
        355                 360                 365

Ile Ser Gln Tyr Leu Gln Ser Thr His Ala Pro Thr His Ser Asp Tyr
    370                 375                 380

Thr Met Thr Leu Leu Asp Leu Phe Glu Val Glu Lys Asp Gly Glu Lys
385                 390                 395                 400

Glu Ala Phe Arg Glu Asp Leu His Asn Arg Met Leu Leu Trp His Gly
                405                 410                 415
```

```
Ser Arg Met Ser Asn Trp Val Gly Ile Leu Ser His Gly Leu Arg Ile
            420                 425                 430

Ala Pro Pro Glu Ala Pro Ile Thr Gly Tyr Met Phe Gly Lys Gly Ile
        435                 440                 445

Tyr Phe Ala Asp Met Ser Ser Lys Ser Ala Asn Tyr Cys Phe Ala Ser
    450                 455                 460

Arg Leu Lys Asn Thr Gly Leu Leu Leu Leu Ser Glu Val Ala Leu Gly
465                 470                 475                 480

Gln Cys Asn Glu Leu Leu Glu Ala Asn Pro Lys Ala Glu Gly Leu Leu
                485                 490                 495

Gln Gly Lys His Ser Thr Lys Gly Leu Gly Lys Met Ala Pro Ser Ser
            500                 505                 510

Ala His Phe Val Thr Leu Asn Gly Ser Thr Val Pro Leu Gly Pro Ala
        515                 520                 525

Ser Asp Thr Gly Ile Leu Asn Pro Asp Gly Tyr Thr Leu Asn Tyr Asn
    530                 535                 540

Glu Tyr Ile Val Tyr Asn Pro Asn Gln Val Arg Met Arg Tyr Leu Leu
545                 550                 555                 560

Lys Val Gln Phe Asn Phe Leu Gln Leu Trp  *
                565                 570
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2265 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(vi) ORIGINAL SOURCE:
(F) TISSUE TYPE: uterus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 242..1843
(D) OTHER INFORMATION:/product= "Poly ADP Ribose Polymerase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
TGGGACTGGT CGCCTGACTC GGCCTGCCCC AGCCTCTGCT TCACCCCACT GGTGGCCAAA      60

TAGCCGATGT CTAATCCCCC ACACAAGCTC ATCCCCGGCC TCTGGGATTG TTGGGAATTC     120

TCTCCCTAAT TCACGCCTGA GGCTCATGGA GAGTTGCTAG ACCTGGGACT GCCCTGGGAG     180

GCGCACACAA CCAGGCCGGG TGGCAGCCAG GACCTCTCCC ATGTCCCTGC TTTTCTTGGC     240

C ATG GCT CCA AAG CCG AAG CCC TGG GTA CAG ACT GAG GGC CCT GAG          286
  Met Ala Pro Lys Pro Lys Pro Trp Val Gln Thr Glu Gly Pro Glu
              575                 580                 585

AAG AAG AAG GGC CGG CAG GCA GGA AGG GAG GAG GAC CCC TTC CGC TCC      334
Lys Lys Lys Gly Arg Gln Ala Gly Arg Glu Glu Asp Pro Phe Arg Ser
            590                 595                 600

ACC GCT GAG GCC CTC AAG GCC ATA CCC GCA GAG AAG CGC ATA ATC CGC      382
Thr Ala Glu Ala Leu Lys Ala Ile Pro Ala Glu Lys Arg Ile Ile Arg
        605                 610                 615

GTG GAT CCA ACA TGT CCA CTC AGC AGC AAC CCC GGG ACC CAG GTG TAT      430
Val Asp Pro Thr Cys Pro Leu Ser Ser Asn Pro Gly Thr Gln Val Tyr
    620                 625                 630

GAG GAC TAC AAC TGC ACC CTG AAC CAG ACC AAC ATC GAG AAC AAC AAC      478
Glu Asp Tyr Asn Cys Thr Leu Asn Gln Thr Asn Ile Glu Asn Asn Asn
635                 640                 645                 650

AAC AAG TTC TAC ATC ATC CAG CTG CTC CAA GAC AGC AAC CGC TTC TTC      526
Asn Lys Phe Tyr Ile Ile Gln Leu Leu Gln Asp Ser Asn Arg Phe Phe
                655                 660                 665

ACC TGC TGG AAC CGC TGG GGC CGT GTG GGA GAG GTC GGC CAG TCA AAG      574
Thr Cys Trp Asn Arg Trp Gly Arg Val Gly Glu Val Gly Gln Ser Lys
            670                 675                 680

ATC AAC CAC TTC ACA AGG CTA GAA GAT GCA AAG AAG GAC TTT GAG AAG      622
Ile Asn His Phe Thr Arg Leu Glu Asp Ala Lys Lys Asp Phe Glu Lys
```

```
        685                 690                 695

AAA TTT CGG GAA AAG ACC AAG AAC AAC TGG GCA GAG CGG GAC CAC TTT      670
Lys Phe Arg Glu Lys Thr Lys Asn Asn Trp Ala Glu Arg Asp His Phe
    700                 705                 710

GTG TCT CAC CCG GGC AAG TAC ACA CTT ATC GAA GTA CAG GCA GAG GAT      718
Val Ser His Pro Gly Lys Tyr Thr Leu Ile Glu Val Gln Ala Glu Asp
715                 720                 725                 730

GAG GCC CAG GAA GCT GTG GTG AAG GTG GAC AGA GGC CCA GTG AGG ACT      766
Glu Ala Gln Glu Ala Val Val Lys Val Asp Arg Gly Pro Val Arg Thr
                735                 740                 745

GTG ACT AAG CGG GTG CAG CCC TGC TCC CTG GAC CCA GCC ACG CAG AAG      814
Val Thr Lys Arg Val Gln Pro Cys Ser Leu Asp Pro Ala Thr Gln Lys
            750                 755                 760

CTC ATC ACT AAC ATC TTC AGC AAG GAG ATG TTC AAG AAC ACC ATG GCC      862
Leu Ile Thr Asn Ile Phe Ser Lys Glu Met Phe Lys Asn Thr Met Ala
        765                 770                 775

CTC ATG GAC CTG GAT GTG AAG AAG ATG CCC CTG GGA AAG CTG AGC AAG      910
Leu Met Asp Leu Asp Val Lys Lys Met Pro Leu Gly Lys Leu Ser Lys
    780                 785                 790

CAA CAG ATT GCA CGG GGT TTC GAG GCC TTG GAG GCG CTG GAG GAG GCC      958
Gln Gln Ile Ala Arg Gly Phe Glu Ala Leu Glu Ala Leu Glu Glu Ala
795                 800                 805                 810

CTG AAA GGC CCC ACG GAT GGT GGC CAA AGC CTG GAG GAG CTG TCC TCA     1006
Leu Lys Gly Pro Thr Asp Gly Gly Gln Ser Leu Glu Glu Leu Ser Ser
                815                 820                 825

CAC TTT TAC ACC GTC ATC CCG CAC AAC TTC GGC CAC AGC CAG CCC CCG     1054
His Phe Tyr Thr Val Ile Pro His Asn Phe Gly His Ser Gln Pro Pro
            830                 835                 840

CCC ATC AAT TCC CCT GAG CTT CTG CAG GCC AAG AAG GAC ATG CTG CTG     1102
Pro Ile Asn Ser Pro Glu Leu Leu Gln Ala Lys Lys Asp Met Leu Leu
        845                 850                 855

GTG CTG GCG GAC ATC GAG CTG GCC CAG GCC CTG CAG GCA GTC TCT GAG     1150
Val Leu Ala Asp Ile Glu Leu Ala Gln Ala Leu Gln Ala Val Ser Glu
    860                 865                 870

CAG GAG AAG ACG GTG GAG GAG GTG CCA CAC CCC CTG GAC CGA GAC TAC     1198
```

```
Gln Glu Lys Thr Val Glu Glu Val Pro His Pro Leu Asp Arg Asp Tyr
875                 880                 885                 890

CAG CTT CTC AAG TGC CAG CTG CAG CTG CTA GAC TCT GGA GCA CCT GAG     1246
Gln Leu Leu Lys Cys Gln Leu Gln Leu Leu Asp Ser Gly Ala Pro Glu
                895                 900                 905

TAC AAG GTG ATA CAG ACC TAC TTA GAA CAG ACT GGC AGC AAC CAC AGG     1294
Tyr Lys Val Ile Gln Thr Tyr Leu Glu Gln Thr Gly Ser Asn His Arg
            910                 915                 920

TGC CCT ACA CTT CAA CAC ATC TGG AAA GTA AAC CAA GAA GGG GAG GAA     1342
Cys Pro Thr Leu Gln His Ile Trp Lys Val Asn Gln Glu Gly Glu Glu
        925                 930                 935

GAC AGA TTC CAG GCC CAC TCC AAA CTG GGT AAT CGG AAG CTG CTG TGG     1390
Asp Arg Phe Gln Ala His Ser Lys Leu Gly Asn Arg Lys Leu Leu Trp
    940                 945                 950

CAT GGC ACC AAC ATG GCC GTG GTG GCC GCC ATC CTC ACT AGT GGG CTC     1438
His Gly Thr Asn Met Ala Val Val Ala Ala Ile Leu Thr Ser Gly Leu
955                 960                 965                 970

CGC ATC ATG CCA CAT TCT GGT GGG CGT GTT GGC AAG GGC ATC TAC TTT     1486
Arg Ile Met Pro His Ser Gly Gly Arg Val Gly Lys Gly Ile Tyr Phe
                975                 980                 985

GCC TCA GAG AAC AGC AAG TCA GCT GGA TAT GTT ATT GGC ATG AAG TGT     1534
Ala Ser Glu Asn Ser Lys Ser Ala Gly Tyr Val Ile Gly Met Lys Cys
            990                 995                1000

GGG GCC CAC CAT GTC GGC TAC ATG TTC CTG GGT GAG GTG GCC CTG GGC     1582
Gly Ala His His Val Gly Tyr Met Phe Leu Gly Glu Val Ala Leu Gly
        1005                1010                1015

AGA GAG CAC CAT ATC AAC ACG GAC AAC CCC AGC TTG AAG AGC CCA CCT     1630
Arg Glu His His Ile Asn Thr Asp Asn Pro Ser Leu Lys Ser Pro Pro
    1020                1025                1030

CCT GGC TTC GAC AGT GTC ATT GCC CGA GGC CAC ACC GAG CCT GAT CCG     1678
Pro Gly Phe Asp Ser Val Ile Ala Arg Gly His Thr Glu Pro Asp Pro
1035                1040                1045                1050

ACC CAG GAC ACT GAG TTG GAG CTG GAT GGC CAG CAA GTG GTG GTG CCC     1726
Thr Gln Asp Thr Glu Leu Glu Leu Asp Gly Gln Gln Val Val Val Pro
                1055                1060                1065
```

```
CAG GGC CAG CCT GTG CCC TGC CCA GAG TTC AGC AGC TCC ACA TTC TCC      1774
Gln Gly Gln Pro Val Pro Cys Pro Glu Phe Ser Ser Ser Thr Phe Ser
            1070                1075                1080

CAG AGC GAG TAC CTC ATC TAC CAG GAG AGC CAG TGT CGC CTG CGC TAC      1822
Gln Ser Glu Tyr Leu Ile Tyr Gln Glu Ser Gln Cys Arg Leu Arg Tyr
        1085                1090                1095

CTG CTG GAG GTC CAC CTC TGA GTGCCCGCCC TGTCCCCCGG GGTCCTGCAA         1873
Leu Leu Glu Val His Leu  *
    1100                1105

GGCTGGACTG TGATCTTCAA TCATCCTGCC CATCTCTGGT ACCCCTATAT CACTCCTTTT    1933

TTTCAAGAAT ACAATACGTT GTTGTTAACT ATAGTCACCA TGCTGTACAA GATCCCTGAA    1993

CTTATGCCTC CTAACTGAAA TTTTGTATTC TTTGACACAT CTGCCCAGTC CCTCTCCTCC    2053

CAGCCCATGG TAACCAGCAT TTGACTCTTT ACTTGTATAA GGGCAGCTTT TATAGGTTCC    2113

ACATGTAAGT GAGATCATGC AGTGTTTGTC TTTCTGTGCC TGGCTTATTT CACTCAGCAT    2173

AATGTGCACC GGGTTCACCC ATGTTTTCAT AAATGACAAG ATTTCCTCCT TTAAAAAAAA    2233

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA                                  2265
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 534 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ala Pro Lys Pro Lys Pro Trp Val Gln Thr Glu Gly Pro Glu Lys
 1               5                  10                  15

Lys Lys Gly Arg Gln Ala Gly Arg Glu Glu Asp Pro Phe Arg Ser Thr
            20                  25                  30

Ala Glu Ala Leu Lys Ala Ile Pro Ala Glu Lys Arg Ile Ile Arg Val
        35                  40                  45

Asp Pro Thr Cys Pro Leu Ser Ser Asn Pro Gly Thr Gln Val Tyr Glu
```

50 55 60

Asp Tyr Asn Cys Thr Leu Asn Gln Thr Asn Ile Glu Asn Asn Asn Asn
65 70 75 80

Lys Phe Tyr Ile Ile Gln Leu Leu Gln Asp Ser Asn Arg Phe Phe Thr
85 90 95

Cys Trp Asn Arg Trp Gly Arg Val Gly Glu Val Gly Gln Ser Lys Ile
100 105 110

Asn His Phe Thr Arg Leu Glu Asp Ala Lys Lys Asp Phe Glu Lys Lys
115 120 125

Phe Arg Glu Lys Thr Lys Asn Asn Trp Ala Glu Arg Asp His Phe Val
130 135 140

Ser His Pro Gly Lys Tyr Thr Leu Ile Glu Val Gln Ala Glu Asp Glu
145 150 155 160

Ala Gln Glu Ala Val Val Lys Val Asp Arg Gly Pro Val Arg Thr Val
165 170 175

Thr Lys Arg Val Gln Pro Cys Ser Leu Asp Pro Ala Thr Gln Lys Leu
180 185 190

Ile Thr Asn Ile Phe Ser Lys Glu Met Phe Lys Asn Thr Met Ala Leu
195 200 205

Met Asp Leu Asp Val Lys Lys Met Pro Leu Gly Lys Leu Ser Lys Gln
210 215 220

Gln Ile Ala Arg Gly Phe Glu Ala Leu Glu Ala Leu Glu Glu Ala Leu
225 230 235 240

Lys Gly Pro Thr Asp Gly Gly Gln Ser Leu Glu Glu Leu Ser Ser His
245 250 255

Phe Tyr Thr Val Ile Pro His Asn Phe Gly His Ser Gln Pro Pro Pro
260 265 270

Ile Asn Ser Pro Glu Leu Leu Gln Ala Lys Lys Asp Met Leu Leu Val
275 280 285

Leu Ala Asp Ile Glu Leu Ala Gln Ala Leu Gln Ala Val Ser Glu Gln
290 295 300

```
Glu Lys Thr Val Glu Glu Val Pro His Pro Leu Asp Arg Asp Tyr Gln
305                 310                 315                 320

Leu Leu Lys Cys Gln Leu Gln Leu Leu Asp Ser Gly Ala Pro Glu Tyr
                325                 330                 335

Lys Val Ile Gln Thr Tyr Leu Glu Gln Thr Gly Ser Asn His Arg Cys
            340                 345                 350

Pro Thr Leu Gln His Ile Trp Lys Val Asn Gln Glu Gly Glu Glu Asp
        355                 360                 365

Arg Phe Gln Ala His Ser Lys Leu Gly Asn Arg Lys Leu Leu Trp His
    370                 375                 380

Gly Thr Asn Met Ala Val Val Ala Ala Ile Leu Thr Ser Gly Leu Arg
385                 390                 395                 400

Ile Met Pro His Ser Gly Gly Arg Val Gly Lys Gly Ile Tyr Phe Ala
                405                 410                 415

Ser Glu Asn Ser Lys Ser Ala Gly Tyr Val Ile Gly Met Lys Cys Gly
            420                 425                 430

Ala His His Val Gly Tyr Met Phe Leu Gly Glu Val Ala Leu Gly Arg
        435                 440                 445

Glu His His Ile Asn Thr Asp Asn Pro Ser Leu Lys Ser Pro Pro Pro
    450                 455                 460

Gly Phe Asp Ser Val Ile Ala Arg Gly His Thr Glu Pro Asp Pro Thr
465                 470                 475                 480

Gln Asp Thr Glu Leu Glu Leu Asp Gly Gln Gln Val Val Val Pro Gln
                485                 490                 495

Gly Gln Pro Val Pro Cys Pro Glu Phe Ser Ser Ser Thr Phe Ser Gln
            500                 505                 510

Ser Glu Tyr Leu Ile Tyr Gln Glu Ser Gln Cys Arg Leu Arg Tyr Leu
        515                 520                 525

Leu Glu Val His Leu  *
    530
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2265 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(vi) ORIGINAL SOURCE:
(F) TISSUE TYPE: uterus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 221..1843
(D) OTHER INFORMATION:/product= "Poly ADP Ribose Polymerase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
TGGGACTGGT CGCCTGACTC GGCCTGCCCC AGCCTCTGCT TCACCCCACT GGTGGCCAAA     60

TAGCCGATGT CTAATCCCCC ACACAAGCTC ATCCCCGGCC TCTGGGATTG TTGGGAATTC    120

TCTCCCTAAT TCACGCCTGA GGCTCATGGA GAGTTGCTAG ACCTGGGACT GCCCTGGGAG    180

GCGCACACAA CCAGGCCGGG TGGCAGCCAG GACCTCTCCC ATG TCC CTG CTT TTC      235
                                           Met Ser Leu Leu Phe
                                           535

TTG GCC ATG GCT CCA AAG CCG AAG CCC TGG GTA CAG ACT GAG GGC CCT      283
Leu Ala Met Ala Pro Lys Pro Lys Pro Trp Val Gln Thr Glu Gly Pro
540                 545                 550                 555

GAG AAG AAG AAG GGC CGG CAG GCA GGA AGG GAG GAG GAC CCC TTC CGC      331
Glu Lys Lys Lys Gly Arg Gln Ala Gly Arg Glu Glu Asp Pro Phe Arg
                560                 565                 570

TCC ACC GCT GAG GCC CTC AAG GCC ATA CCC GCA GAG AAG CGC ATA ATC      379
Ser Thr Ala Glu Ala Leu Lys Ala Ile Pro Ala Glu Lys Arg Ile Ile
            575                 580                 585

CGC GTG GAT CCA ACA TGT CCA CTC AGC AGC AAC CCC GGG ACC CAG GTG      427
Arg Val Asp Pro Thr Cys Pro Leu Ser Ser Asn Pro Gly Thr Gln Val
```

590 595 600

```
TAT GAG GAC TAC AAC TGC ACC CTG AAC CAG ACC AAC ATC GAG AAC AAC      475
Tyr Glu Asp Tyr Asn Cys Thr Leu Asn Gln Thr Asn Ile Glu Asn Asn
    605                 610                 615

AAC AAC AAG TTC TAC ATC ATC CAG CTG CTC CAA GAC AGC AAC CGC TTC      523
Asn Asn Lys Phe Tyr Ile Ile Gln Leu Leu Gln Asp Ser Asn Arg Phe
620                 625                 630                 635

TTC ACC TGC TGG AAC CGC TGG GGC CGT GTG GGA GAG GTC GGC CAG TCA      571
Phe Thr Cys Trp Asn Arg Trp Gly Arg Val Gly Glu Val Gly Gln Ser
                640                 645                 650

AAG ATC AAC CAC TTC ACA AGG CTA GAA GAT GCA AAG AAG GAC TTT GAG      619
Lys Ile Asn His Phe Thr Arg Leu Glu Asp Ala Lys Lys Asp Phe Glu
            655                 660                 665

AAG AAA TTT CGG GAA AAG ACC AAG AAC AAC TGG GCA GAG CGG GAC CAC      667
Lys Lys Phe Arg Glu Lys Thr Lys Asn Asn Trp Ala Glu Arg Asp His
        670                 675                 680

TTT GTG TCT CAC CCG GGC AAG TAC ACA CTT ATC GAA GTA CAG GCA GAG      715
Phe Val Ser His Pro Gly Lys Tyr Thr Leu Ile Glu Val Gln Ala Glu
    685                 690                 695

GAT GAG GCC CAG GAA GCT GTG GTG AAG GTG GAC AGA GGC CCA GTG AGG      763
Asp Glu Ala Gln Glu Ala Val Val Lys Val Asp Arg Gly Pro Val Arg
700                 705                 710                 715

ACT GTG ACT AAG CGG GTG CAG CCC TGC TCC CTG GAC CCA GCC ACG CAG      811
Thr Val Thr Lys Arg Val Gln Pro Cys Ser Leu Asp Pro Ala Thr Gln
                720                 725                 730

AAG CTC ATC ACT AAC ATC TTC AGC AAG GAG ATG TTC AAG AAC ACC ATG      859
Lys Leu Ile Thr Asn Ile Phe Ser Lys Glu Met Phe Lys Asn Thr Met
            735                 740                 745

GCC CTC ATG GAC CTG GAT GTG AAG AAG ATG CCC CTG GGA AAG CTG AGC      907
Ala Leu Met Asp Leu Asp Val Lys Lys Met Pro Leu Gly Lys Leu Ser
        750                 755                 760

AAG CAA CAG ATT GCA CGG GGT TTC GAG GCC TTG GAG GCG CTG GAG GAG      955
Lys Gln Gln Ile Ala Arg Gly Phe Glu Ala Leu Glu Ala Leu Glu Glu
    765                 770                 775

GCC CTG AAA GGC CCC ACG GAT GGT GGC CAA AGC CTG GAG GAG CTG TCC     1003
```

```
Ala Leu Lys Gly Pro Thr Asp Gly Gly Gln Ser Leu Glu Glu Leu Ser
780                 785                 790                 795

TCA CAC TTT TAC ACC GTC ATC CCG CAC AAC TTC GGC CAC AGC CAG CCC     1051
Ser His Phe Tyr Thr Val Ile Pro His Asn Phe Gly His Ser Gln Pro
                800                 805                 810

CCG CCC ATC AAT TCC CCT GAG CTT CTG CAG GCC AAG AAG GAC ATG CTG     1099
Pro Pro Ile Asn Ser Pro Glu Leu Leu Gln Ala Lys Lys Asp Met Leu
            815                 820                 825

CTG GTG CTG GCG GAC ATC GAG CTG GCC CAG GCC CTG CAG GCA GTC TCT     1147
Leu Val Leu Ala Asp Ile Glu Leu Ala Gln Ala Leu Gln Ala Val Ser
        830                 835                 840

GAG CAG GAG AAG ACG GTG GAG GAG GTG CCA CAC CCC CTG GAC CGA GAC     1195
Glu Gln Glu Lys Thr Val Glu Glu Val Pro His Pro Leu Asp Arg Asp
    845                 850                 855

TAC CAG CTT CTC AAG TGC CAG CTG CAG CTG CTA GAC TCT GGA GCA CCT     1243
Tyr Gln Leu Leu Lys Cys Gln Leu Gln Leu Leu Asp Ser Gly Ala Pro
860                 865                 870                 875

GAG TAC AAG GTG ATA CAG ACC TAC TTA GAA CAG ACT GGC AGC AAC CAC     1291
Glu Tyr Lys Val Ile Gln Thr Tyr Leu Glu Gln Thr Gly Ser Asn His
                880                 885                 890

AGG TGC CCT ACA CTT CAA CAC ATC TGG AAA GTA AAC CAA GAA GGG GAG     1339
Arg Cys Pro Thr Leu Gln His Ile Trp Lys Val Asn Gln Glu Gly Glu
            895                 900                 905

GAA GAC AGA TTC CAG GCC CAC TCC AAA CTG GGT AAT CGG AAG CTG CTG     1387
Glu Asp Arg Phe Gln Ala His Ser Lys Leu Gly Asn Arg Lys Leu Leu
        910                 915                 920

TGG CAT GGC ACC AAC ATG GCC GTG GTG GCC GCC ATC CTC ACT AGT GGG     1435
Trp His Gly Thr Asn Met Ala Val Val Ala Ala Ile Leu Thr Ser Gly
    925                 930                 935

CTC CGC ATC ATG CCA CAT TCT GGT GGG CGT GTT GGC AAG GGC ATC TAC     1483
Leu Arg Ile Met Pro His Ser Gly Gly Arg Val Gly Lys Gly Ile Tyr
940                 945                 950                 955

TTT GCC TCA GAG AAC AGC AAG TCA GCT GGA TAT GTT ATT GGC ATG AAG     1531
Phe Ala Ser Glu Asn Ser Lys Ser Ala Gly Tyr Val Ile Gly Met Lys
                960                 965                 970
```

```
TGT GGG GCC CAC CAT GTC GGC TAC ATG TTC CTG GGT GAG GTG GCC CTG      1579
Cys Gly Ala His His Val Gly Tyr Met Phe Leu Gly Glu Val Ala Leu
            975                 980                 985

GGC AGA GAG CAC CAT ATC AAC ACG GAC AAC CCC AGC TTG AAG AGC CCA      1627
Gly Arg Glu His His Ile Asn Thr Asp Asn Pro Ser Leu Lys Ser Pro
        990                 995                 1000

CCT CCT GGC TTC GAC AGT GTC ATT GCC CGA GGC CAC ACC GAG CCT GAT      1675
Pro Pro Gly Phe Asp Ser Val Ile Ala Arg Gly His Thr Glu Pro Asp
    1005                1010                1015

CCG ACC CAG GAC ACT GAG TTG GAG CTG GAT GGC CAG CAA GTG GTG GTG      1723
Pro Thr Gln Asp Thr Glu Leu Glu Leu Asp Gly Gln Gln Val Val Val
1020                1025                1030                1035

CCC CAG GGC CAG CCT GTG CCC TGC CCA GAG TTC AGC AGC TCC ACA TTC      1771
Pro Gln Gly Gln Pro Val Pro Cys Pro Glu Phe Ser Ser Ser Thr Phe
                1040                1045                1050

TCC CAG AGC GAG TAC CTC ATC TAC CAG GAG AGC CAG TGT CGC CTG CGC      1819
Ser Gln Ser Glu Tyr Leu Ile Tyr Gln Glu Ser Gln Cys Arg Leu Arg
            1055                1060                1065

TAC CTG CTG GAG GTC CAC CTC TGA GTGCCCGCCC TGTCCCCCGG GGTCCTGCAA     1873
Tyr Leu Leu Glu Val His Leu  *
        1070                1075

GGCTGGACTG TGATCTTCAA TCATCCTGCC CATCTCTGGT ACCCCTATAT CACTCCTTTT    1933

TTTCAAGAAT ACAATACGTT GTTGTTAACT ATAGTCACCA TGCTGTACAA GATCCCTGAA    1993

CTTATGCCTC CTAACTGAAA TTTTGTATTC TTTGACACAT CTGCCCAGTC CCTCTCCTCC    2053

CAGCCCATGG TAACCAGCAT TTGACTCTTT ACTTGTATAA GGGCAGCTTT TATAGGTTCC    2113

ACATGTAAGT GAGATCATGC AGTGTTTGTC TTTCTGTGCC TGGCTTATTT CACTCAGCAT    2173

AATGTGCACC GGGTTCACCC ATGTTTTCAT AAATGACAAG ATTTCCTCCT TTAAAAAAAA    2233

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA                                  2265
```

(2) INFORMATION FOR ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 541 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ser Leu Leu Phe Leu Ala Met Ala Pro Lys Pro Lys Pro Trp Val
 1               5                  10                  15

Gln Thr Glu Gly Pro Glu Lys Lys Lys Gly Arg Gln Ala Gly Arg Glu
            20                  25                  30

Glu Asp Pro Phe Arg Ser Thr Ala Glu Ala Leu Lys Ala Ile Pro Ala
        35                  40                  45

Glu Lys Arg Ile Ile Arg Val Asp Pro Thr Cys Pro Leu Ser Ser Asn
    50                  55                  60

Pro Gly Thr Gln Val Tyr Glu Asp Tyr Asn Cys Thr Leu Asn Gln Thr
65                  70                  75                  80

Asn Ile Glu Asn Asn Asn Asn Lys Phe Tyr Ile Ile Gln Leu Leu Gln
                85                  90                  95

Asp Ser Asn Arg Phe Phe Thr Cys Trp Asn Arg Trp Gly Arg Val Gly
            100                 105                 110

Glu Val Gly Gln Ser Lys Ile Asn His Phe Thr Arg Leu Glu Asp Ala
        115                 120                 125

Lys Lys Asp Phe Glu Lys Lys Phe Arg Glu Lys Thr Lys Asn Asn Trp
    130                 135                 140

Ala Glu Arg Asp His Phe Val Ser His Pro Gly Lys Tyr Thr Leu Ile
145                 150                 155                 160

Glu Val Gln Ala Glu Asp Glu Ala Gln Glu Ala Val Val Lys Val Asp
                165                 170                 175

Arg Gly Pro Val Arg Thr Val Thr Lys Arg Val Gln Pro Cys Ser Leu
            180                 185                 190

Asp Pro Ala Thr Gln Lys Leu Ile Thr Asn Ile Phe Ser Lys Glu Met
        195                 200                 205

Phe Lys Asn Thr Met Ala Leu Met Asp Leu Asp Val Lys Lys Met Pro
    210                 215                 220
```

Leu Gly Lys Leu Ser Lys Gln Gln Ile Ala Arg Gly Phe Glu Ala Leu
225 230 235 240

Glu Ala Leu Glu Glu Ala Leu Lys Gly Pro Thr Asp Gly Gly Gln Ser
245 250 255

Leu Glu Glu Leu Ser Ser His Phe Tyr Thr Val Ile Pro His Asn Phe
260 265 270

Gly His Ser Gln Pro Pro Pro Ile Asn Ser Pro Glu Leu Leu Gln Ala
275 280 285

Lys Lys Asp Met Leu Leu Val Leu Ala Asp Ile Glu Leu Ala Gln Ala
290 295 300

Leu Gln Ala Val Ser Glu Gln Glu Lys Thr Val Glu Glu Val Pro His
305 310 315 320

Pro Leu Asp Arg Asp Tyr Gln Leu Leu Lys Cys Gln Leu Gln Leu Leu
325 330 335

Asp Ser Gly Ala Pro Glu Tyr Lys Val Ile Gln Thr Tyr Leu Glu Gln
340 345 350

Thr Gly Ser Asn His Arg Cys Pro Thr Leu Gln His Ile Trp Lys Val
355 360 365

Asn Gln Glu Gly Glu Glu Asp Arg Phe Gln Ala His Ser Lys Leu Gly
370 375 380

Asn Arg Lys Leu Leu Trp His Gly Thr Asn Met Ala Val Val Ala Ala
385 390 395 400

Ile Leu Thr Ser Gly Leu Arg Ile Met Pro His Ser Gly Gly Arg Val
405 410 415

Gly Lys Gly Ile Tyr Phe Ala Ser Glu Asn Ser Lys Ser Ala Gly Tyr
420 425 430

Val Ile Gly Met Lys Cys Gly Ala His His Val Gly Tyr Met Phe Leu
435 440 445

Gly Glu Val Ala Leu Gly Arg Glu His His Ile Asn Thr Asp Asn Pro
450 455 460

Ser Leu Lys Ser Pro Pro Pro Gly Phe Asp Ser Val Ile Ala Arg Gly

```
465                 470                 475                 480

His Thr Glu Pro Asp Pro Thr Gln Asp Thr Glu Leu Glu Leu Asp Gly
                485                 490                 495

Gln Gln Val Val Val Pro Gln Gly Gln Pro Val Pro Cys Pro Glu Phe
            500                 505                 510

Ser Ser Ser Thr Phe Ser Gln Ser Glu Tyr Leu Ile Tyr Gln Glu Ser
        515                 520                 525

Gln Cys Arg Leu Arg Tyr Leu Leu Glu Val His Leu  *
    530                 535                 540
```

We claim:

1. A poly(ADP-ribose) polymerase (PARP) homolog which has an amino acid sequence which has
   a) a functional NAD^+ binding domain and
   b) no zinc finger sequence motif of the general formula

$$CX_2CX_mHX_2C$$

   in which
   m is an integral value from 28 or 30, and the X radicals are, independently of one another, any amino acid;
   and the functional equivalents thereof.

2. A PARP homolog as claimed in claim 1, wherein the functional NAD^+ binding domain comprises the following general sequence motif:

$$LLWHG(S/T)X_7IL(S/T)XGLRIXPX_n(S/T)GX_3GKGIYFAX_3SKSAXY$$

   in which
   n is an integral value from 1 to 5, and the X radicals are, independently of one another, any amino acid.

3. A PARP homolog as claimed in either of the preceding claims, comprising at least one other of the following part-sequence motifs:

$$LX_9NX_2YX_2QLLXDX_{10/11}WGRVG,$$
$$AX_3FXKX_4KTXNXWX_5FX_3PXK,$$
$$QXLIX_2IX_9MX_{10}PLGKLX_3QIX_6L,$$
$$FYTXIPHXFGX_3PP;\ \text{and}$$
$$KX_3LX_2LXDIEXAX_2L,$$

   in which the X radicals are, independently of one another, any amino acid.

4. A human PARP homolog as claimed in any of the preceding claims, which has the amino acid sequence shown in SEQ ID NO: 2 (humanPARP2) or SEQ ID NO: 4 or 6 (humanPARP3 type 1 or 2), and the functional equivalents thereof.

5. A binding partner for PARP homologs as claimed in any of the preceding claims, selected from
   a) antibodies and fragments thereof,

b) protein-like compounds which interact with a part-sequence of the protein, and

c) low molecular weight effectors which modulate the catalytic PARP activity or another biological function of a PARP molecule.

6. A nucleic acid comprising

a) a nucleotide sequence coding for at least one PARP homolog as claimed in any of claims 1 to 4, or the complementary nucleotide sequence thereof;

b) a nucleotide sequence which hybridizes with a sequence as specified in a) under stringent conditions; or

c) nucleotide sequences which are derived from the nucleotide sequences defined in a) and b) through the degeneracy of the genetic code.

7. A nucleic acid as claimed in claim 6, comprising

a) nucleotides +3 to +1715 shown in SEQ ID NO:1;

b) nucleotides +242 to +1843 shown in SEQ ID NO:3; or

c) nucleotides +221 to +1843 shown in SEQ ID NO:5.

8. An expression cassette comprising, under the genetic control of at least regulatory nucleotide sequence, at least one nucleotide sequence as claimed in either of claims 6 and 7.

9. A recombinant vector comprising at least one expression cassette as claimed in claim 8.

10. A recombinant microorganism comprising at least one recombinant vector as claimed in claim 9.

11. A transgenic mammal comprising a vector as claimed in claim 9.

12. A PARP-deficient mammal or PARP-deficient eukaryotic cell, in which functional expression of at least one gene which codes for a PARP homolog as claimed in any of claims 1 to 4 is inhibited.

13. An in vitro screening method for binding partners for a PARP homolog as claimed in any of claims 1 to 4, which comprises

a1) immobilizing at least one PARP homolog on a support;

b1) contacting the immobilized PARP homolog with an analyte in which at least one binding partner is suspected; and

c1) determining, where appropriate after an incubation period, analyte constituents bound to the immobilized PARP homolog;

or

a2) immobilizing an analyte which comprises at least one possible binding partner for the PARP homolog on a support;

b2) contacting the immobilized analyte with at least one PARP homolog for which a binding partner is sought; and

c3) examining the immobilized analyte, where appropriate after an incubation period, for binding of the PARP homolog.

14. A method for the qualitative or quantitative determination of nucleic acids encoding a PARP homolog as claimed in any of claims 1 to 4, which comprises

a) incubating a biological sample with a defined amount of an exogenous nucleic acid as claimed in either of claims 6 and 7, hybridizing under stringent conditions, determining the hybridizing nucleic acids and, where appropriate, comparing with a standard; or

b) incubating a biological sample with a pair of oligonucleotide primers with specificity for a PARP homolog-encoding nucleic acid, amplifying the nucleic acid, determining the amplification product and, where appropriate, comparing with a standard.

15. A method for the qualitative or quantitative determination of a PARP homolog as claimed in any of claims 1 to 4, which comprises

a) incubating a biological sample with a binding partner specific for a PARP homolog,

b) detecting the binding partner/PARP complex and, where appropriate,

c) comparing the result with a standard.

16. A method as claimed in claim 15, wherein the binding partner is an antibody or a binding fragment thereof, which carries a detectable label where appropriate.

17. A method as claimed in any of claims 14 to 16 for diagnosing sepsis- or ischemia-related tissue damage, in particular strokes, myocardial infarcts or septic shock.

18. A method for determining the efficacy of PARP effectors, which comprises
    a) incubating a PARP homolog as claimed in any of claims 1 to 4 with an analyte which comprises an effector of a physiological or pathological PARP activity; removing the effector again where appropriate; and
    b) determining the activity of the PARP homolog, where appropriate after adding substrates or cosubstrates.

19. A composition for gene therapy, which comprises in a vehicle acceptable for gene therapy a nucleic acid construct which
    a) comprises an antisense nucleic acid against a coding nucleic acid as claimed in either of claims 6 and 7; or
    b) a ribozyme against a nucleic acid as claimed in either of claims 6 and 7; or
    c) codes for a specific PARP inhibitor.

20. A pharmaceutical composition comprising, in a pharmaceutically acceptable vehicle, at least one PARP protein as claimed in any of claims 1 to 4, at least one PARP binding partner as claimed in claim 5 or at least one coding nucleotide sequence as claimed in claim 6 or 7.

21. The use of low molecular weight PARP binding partners as claimed in claim 5 for the diagnosis or therapy of pathological states in the development and/or progress of which at least one one PARP protein, or a polypeptide derived therefrom, are [sic] involved.

58/iT

Abstract

The invention relates to poly(ADP-ribose)polymerase (PARP) homologs which have an amino acid sequence which has

a) A functional NAD^+ binding domain
and
b) no zinc finger sequence motif of the general formula

$$CX_2CX_mHX_2C$$

in which
m is an integral value from 28 or 30, and the X radicals are, independently of one another, any amino acid;

and the functional equivalents thereof; nucleic acids coding therefor; antibodies with specificity for the novel protein; pharmaceutical compositions and compositions for gene therapy which comprise products according to the invention; methods for the analytical determination of the proteins and nucleic acids according to the invention; methods for identifying effectors or binding partners of the proteins according to the invention; and methods for determining the efficacy of such effectors.

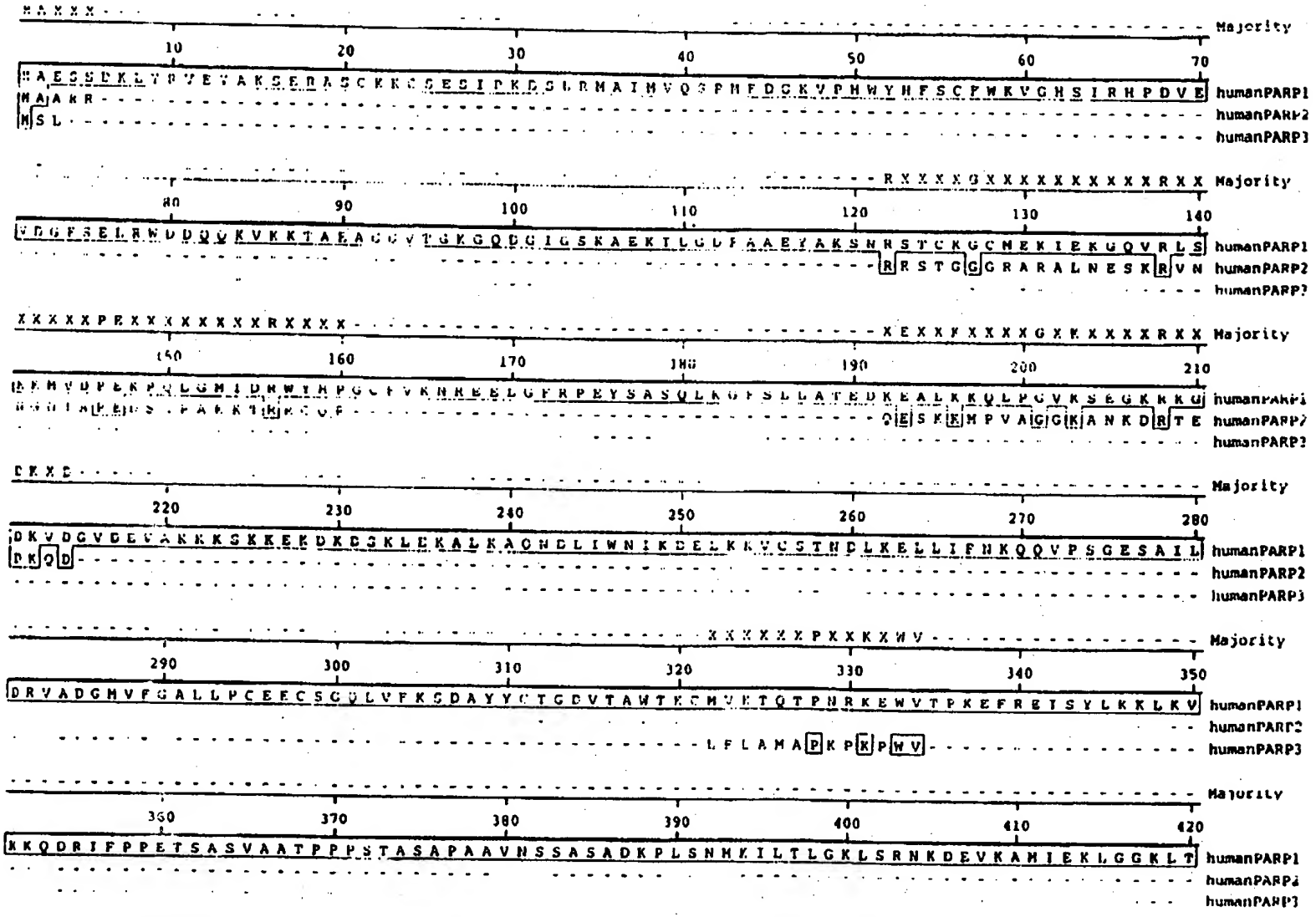


Fig. 1 (1)

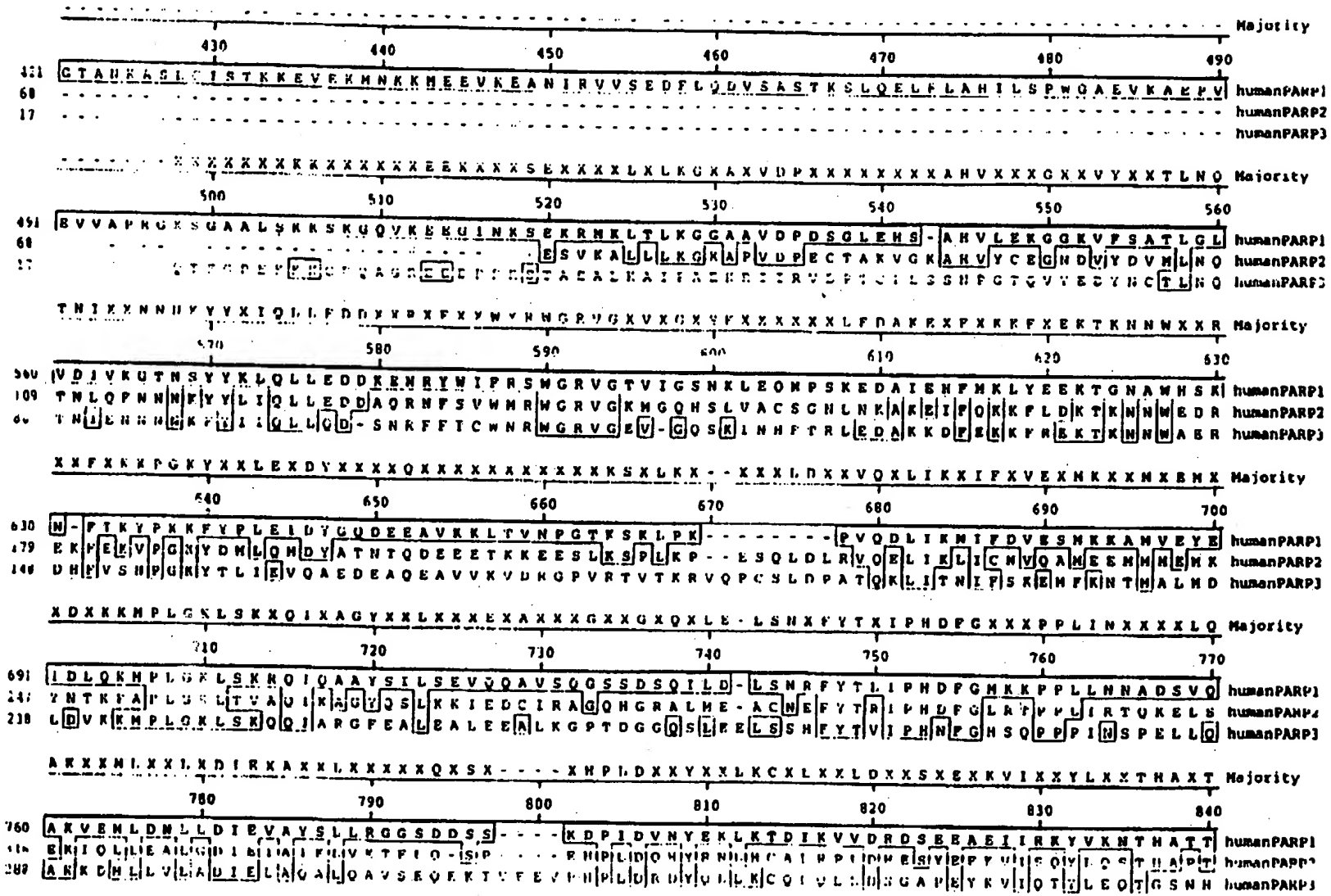


Fig. 1 (2)

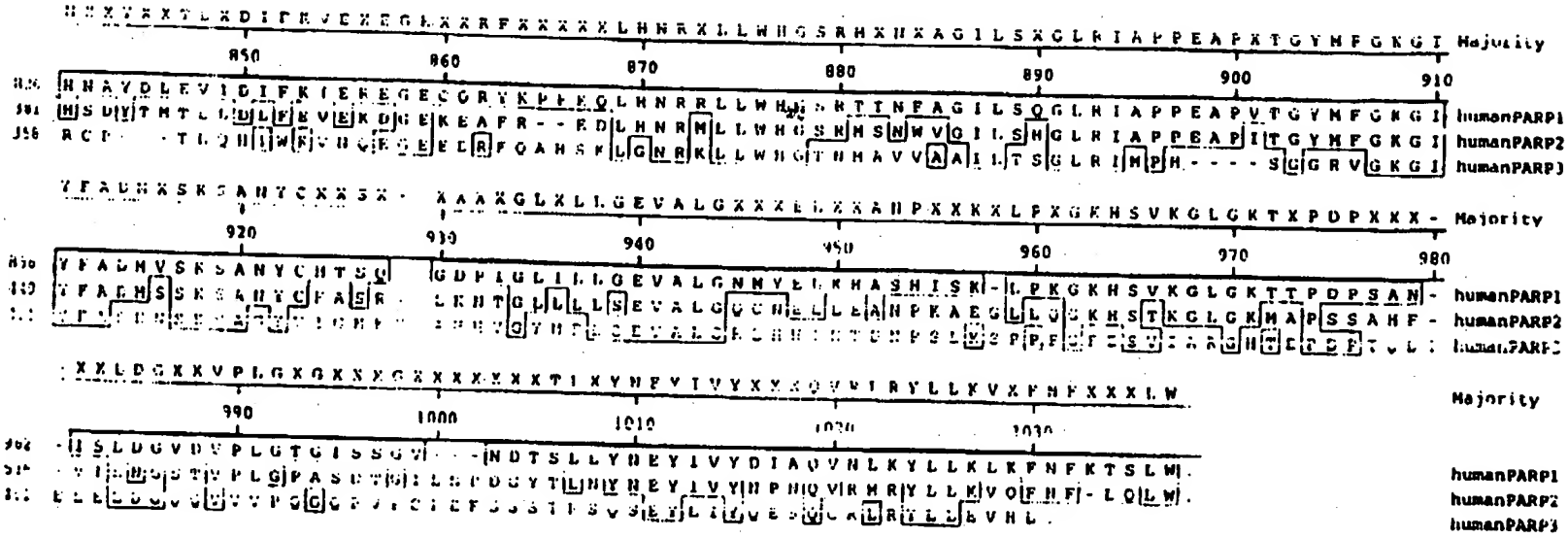
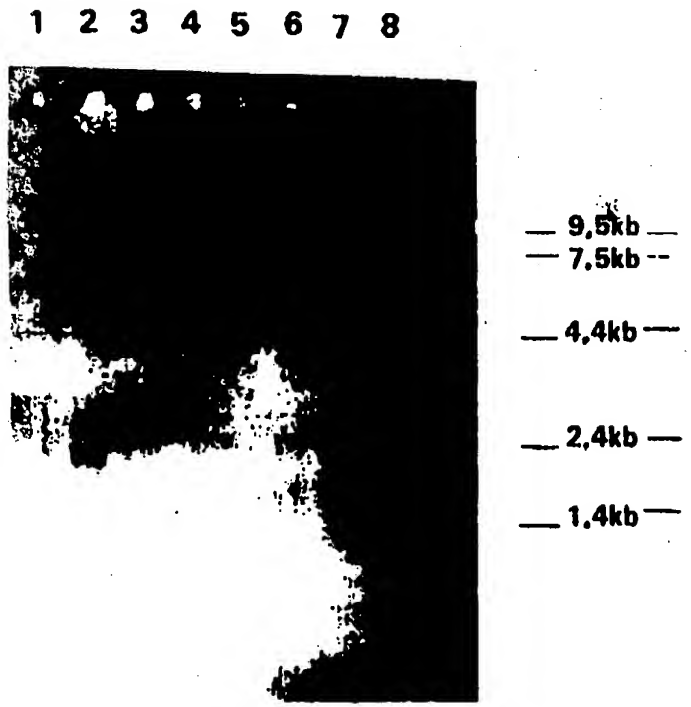


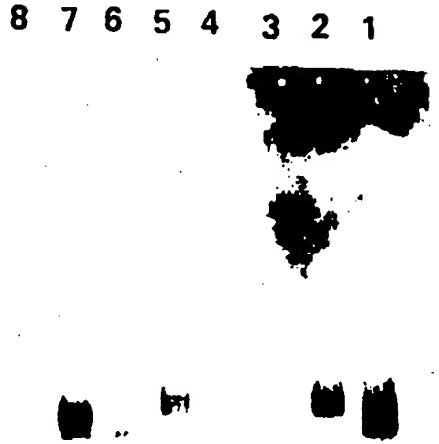
Fig. 1 (3)

Fig. 2



PARP2

(A)



PARP3

(B)